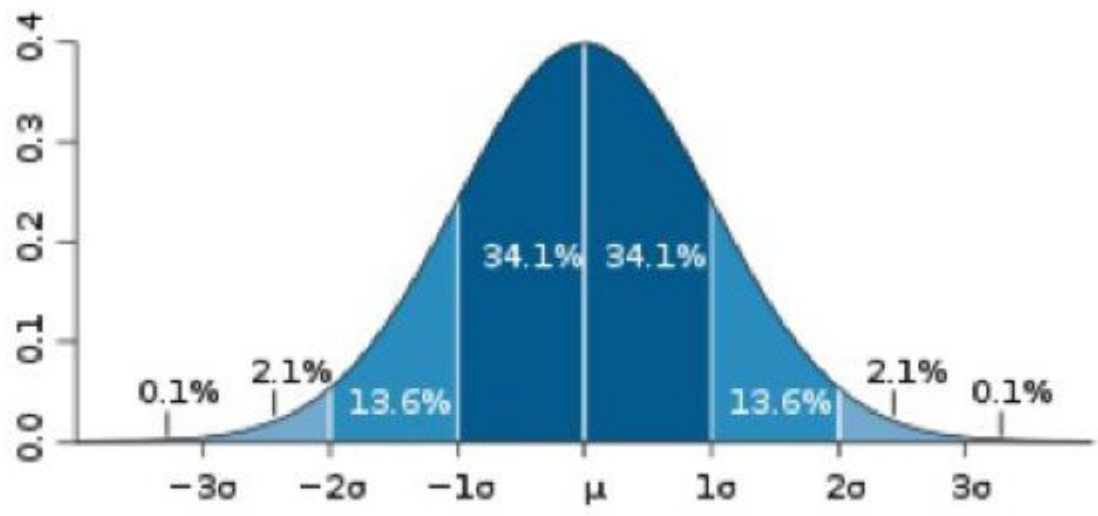
Leis de Potência

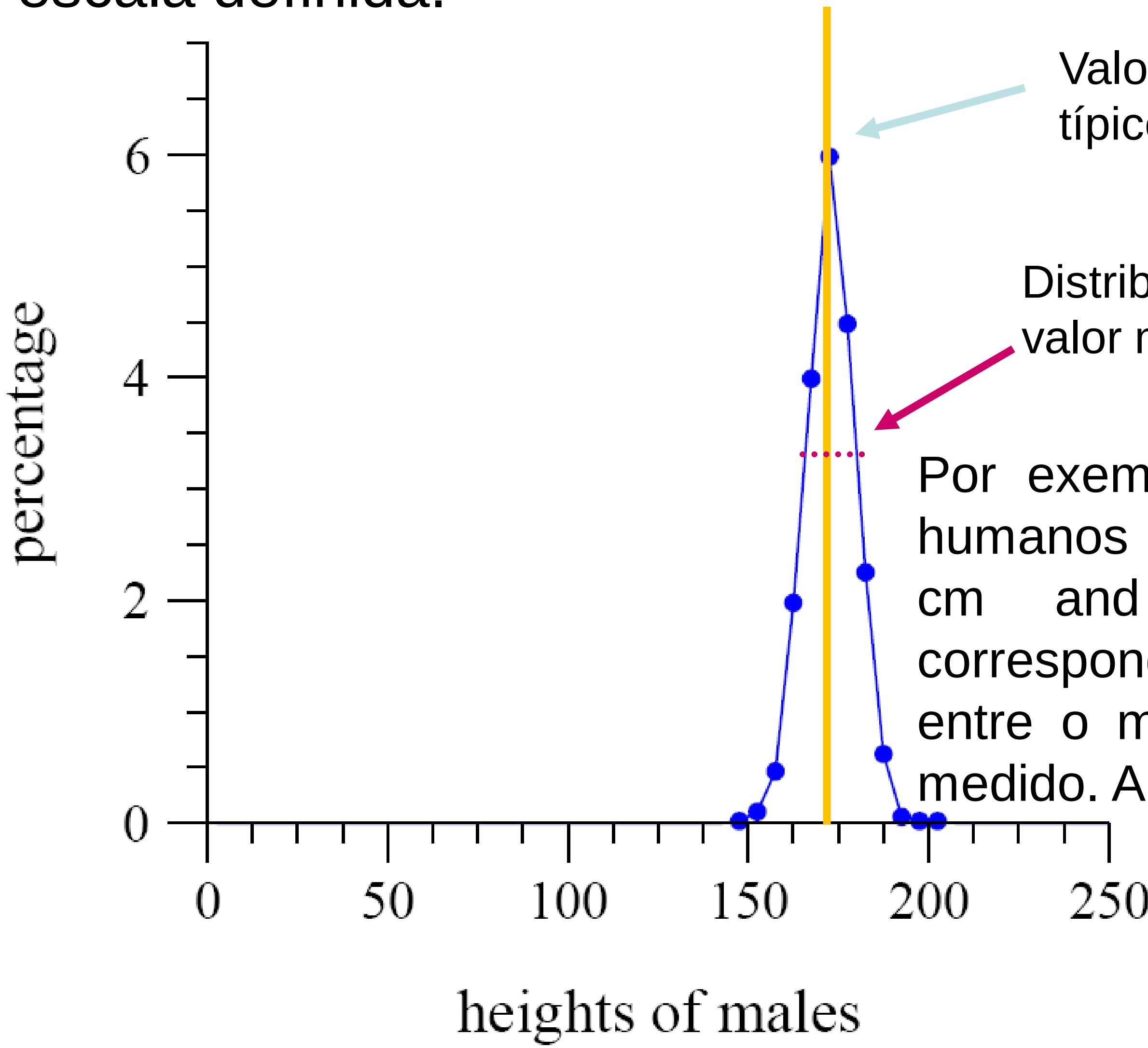
Leis de potência e não-normalidade

- A base das ciências aplicadas é a distribuição normal.
- *O conceito de média.*



Escalas Típicas

Muitas coisas que os cientistas medem tem um tamanho típico ou escala definida.



Valor médio próximo ao mais típico

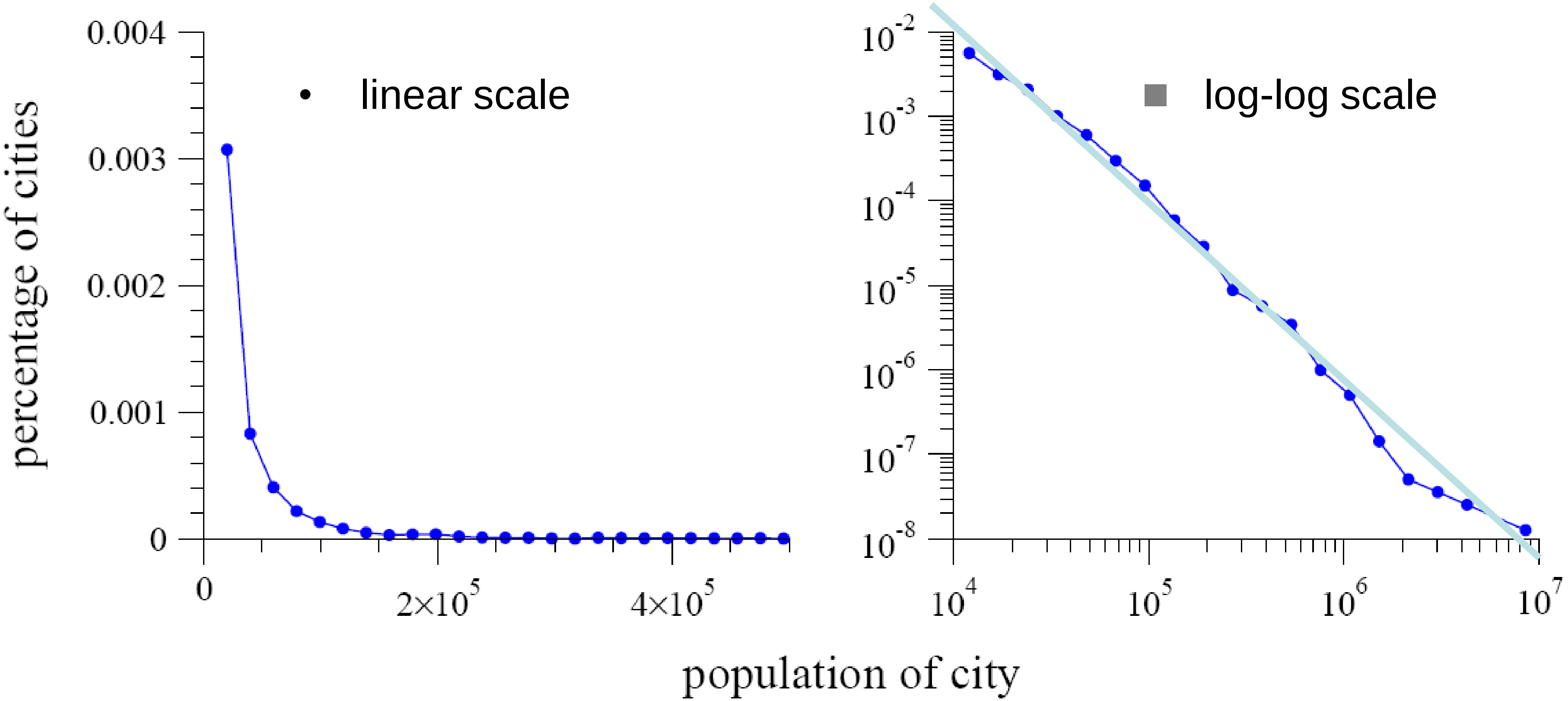
Distributição fica ao redor do valor médio de forma simétrica

Por exemplo, a altura de seres humanos adultos variam entre 50 cm and 272 cm, o que corresponde a uma razão de 4.8 entre o menor e o maior ser já medido. A média é de 175 cm.

Power Law Distribution

$$p(x) = Cx^{-\alpha}$$

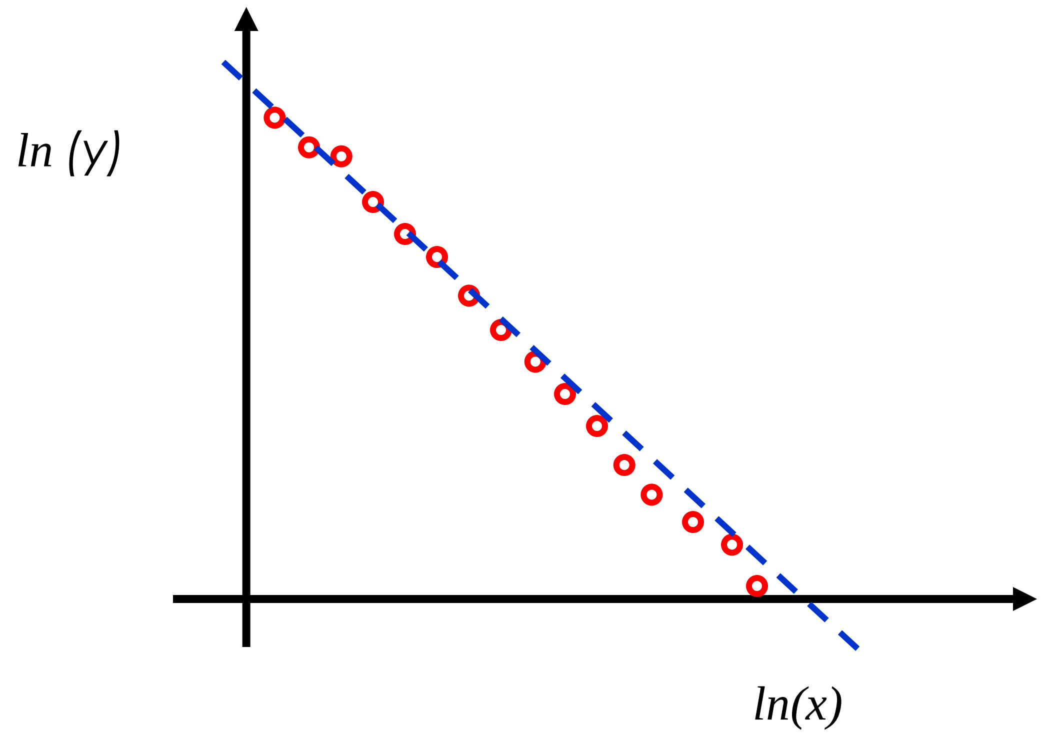
Power-law distribution



- Alta assimetria (asymmetry)
- Linha reda no log-log plot

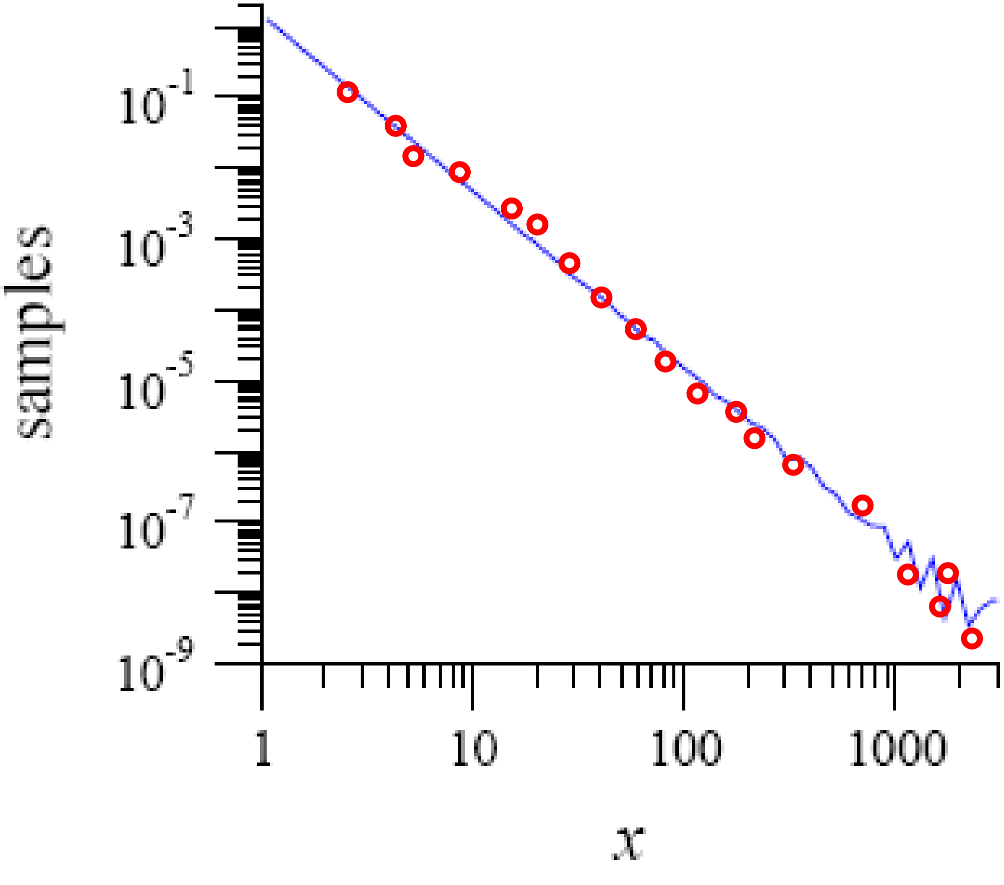
Log-log plot

$$ln(y) = A\, ln(x) + c$$



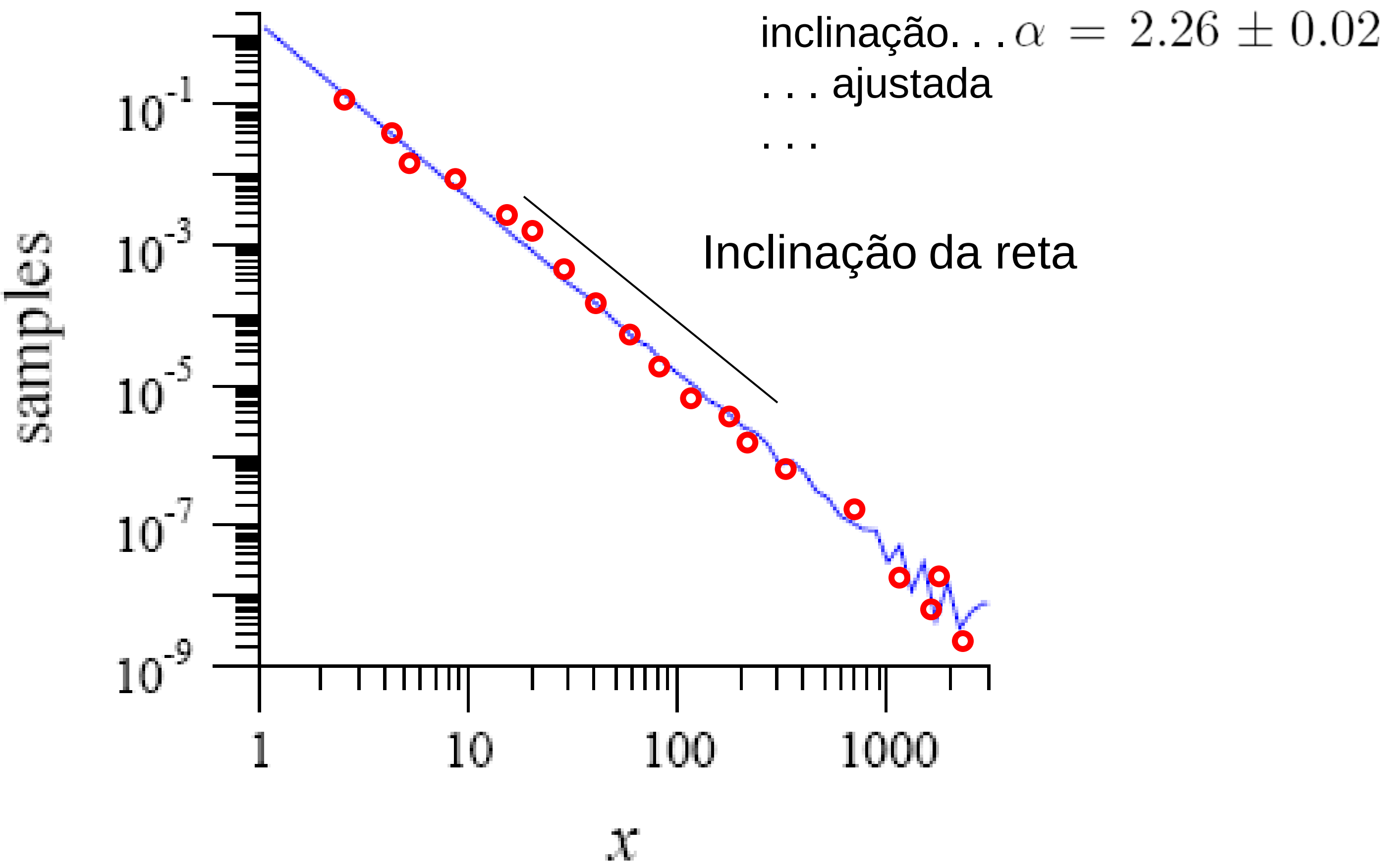
y=# vezes x ocorre

Log-log plot



<u>NO typical value or a typical scale</u> (all sizes, all scales).

Log-log plot



<u>NO typical value or a typical scale</u> (all sizes, all scales).

Leis de potência - exemplos

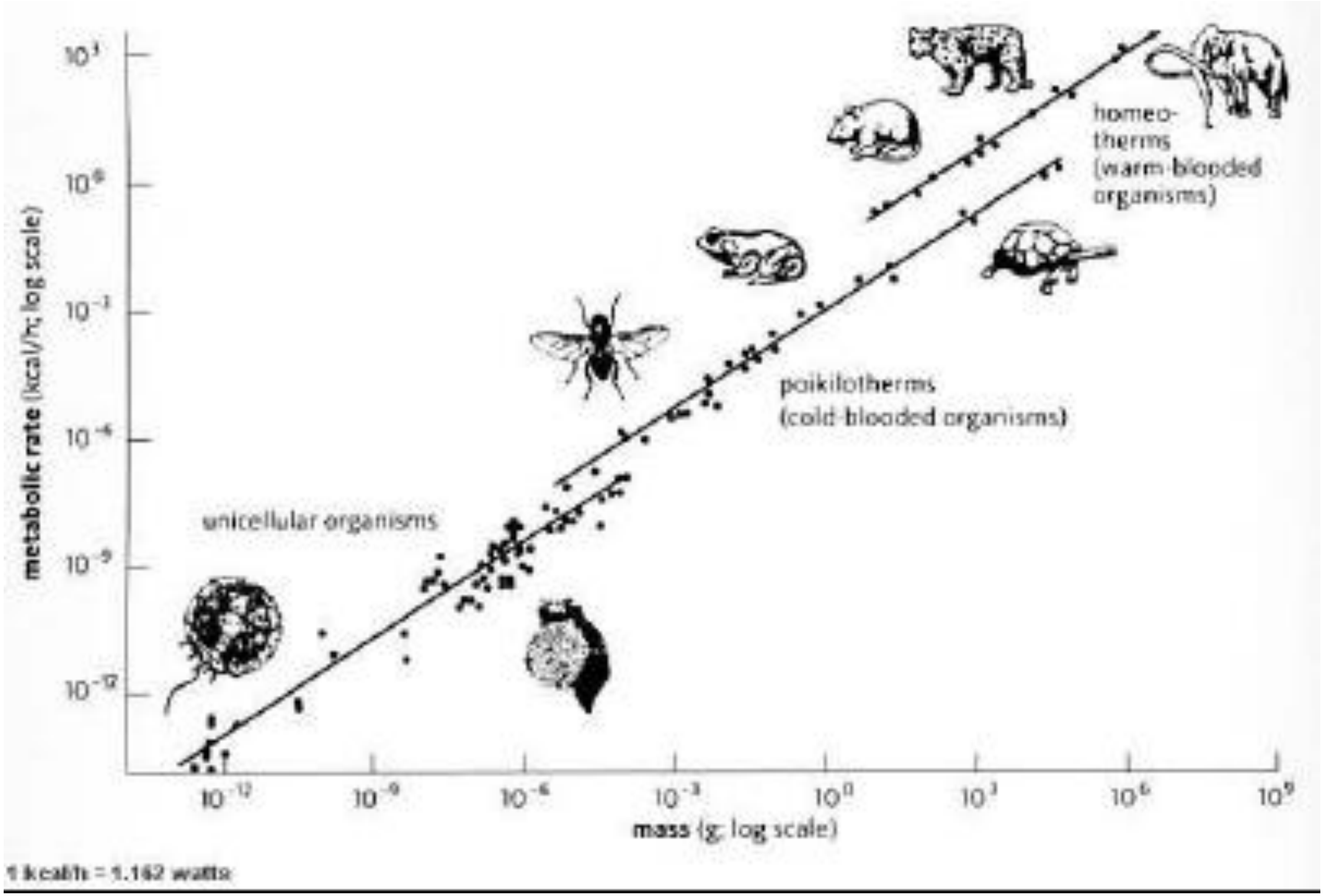
power law exponent α

$$p(x) = Cx^{-\alpha}$$

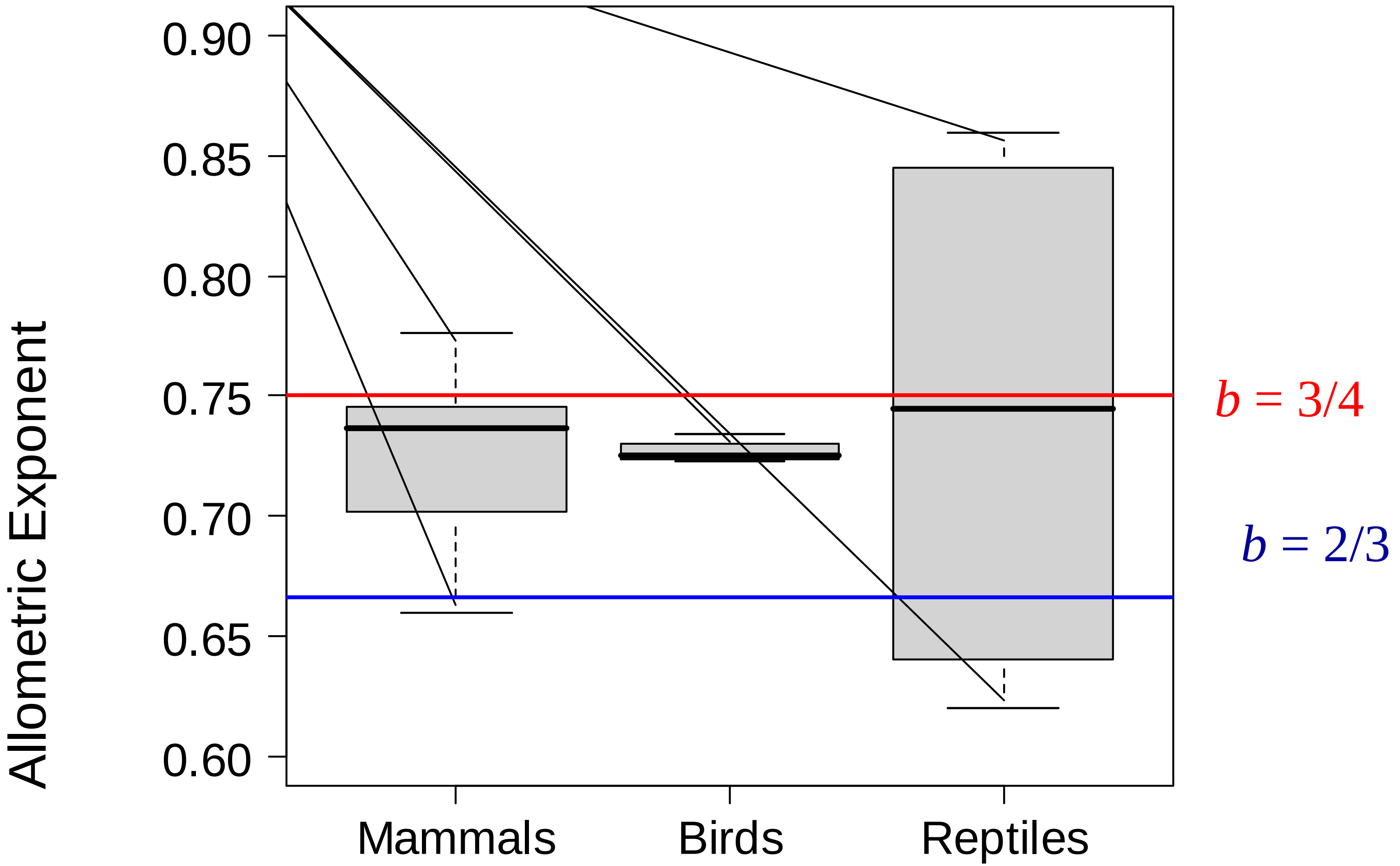
normalization
constant (probabilities over
all *x* must sum to 1)

• Lei de Kleiber $Mt(m) \propto m^{\frac{3}{4}}$

- *Gato tem 100*
- *vezes o peso*
- *do rato e 31*
- *vezes a taxa*
- *metabólica.*

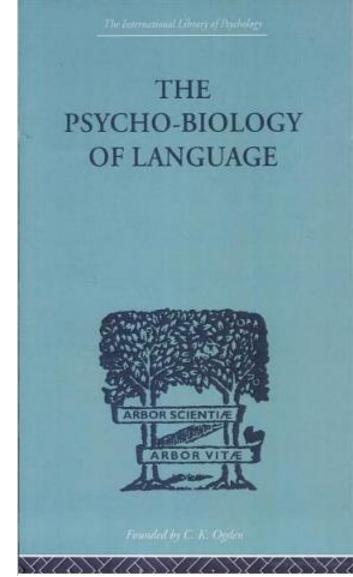


Question 4: estimates often < 3/4?

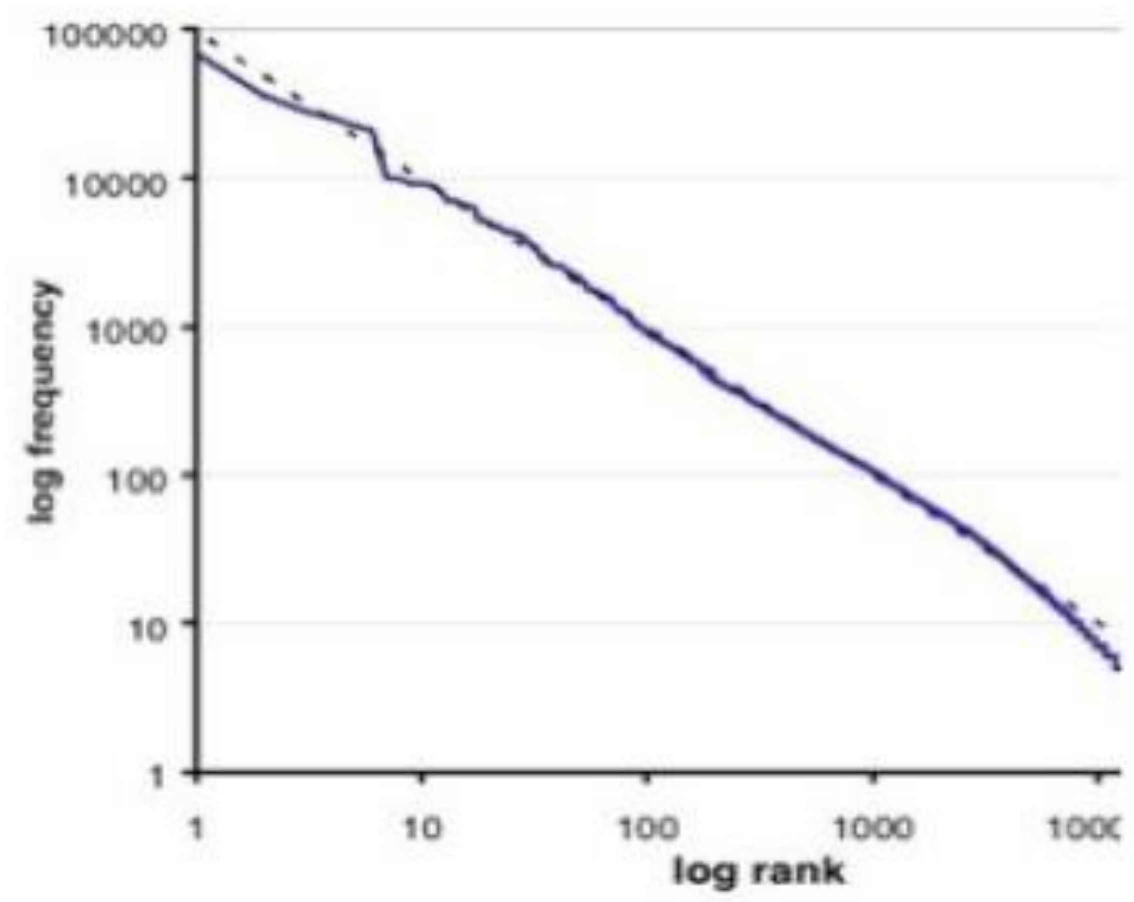


- **Lei de Zipf**

$$f(x) = x^{-1}$$



- *A segunda palavra no raking (x) tem a metade da probabilidade de ocorrência que a primeira.*



- **Lei de Pareto**

- The Italian economist Vilfredo Pareto was interested in the distribution of income.
- Pareto's law is expressed in terms of the cumulative distribution
  - the probability that a person earns X or more

$$P[X > x] \sim x^{-k}$$

- Here we recognize *k* as just α -1, where α is the power-law exponent

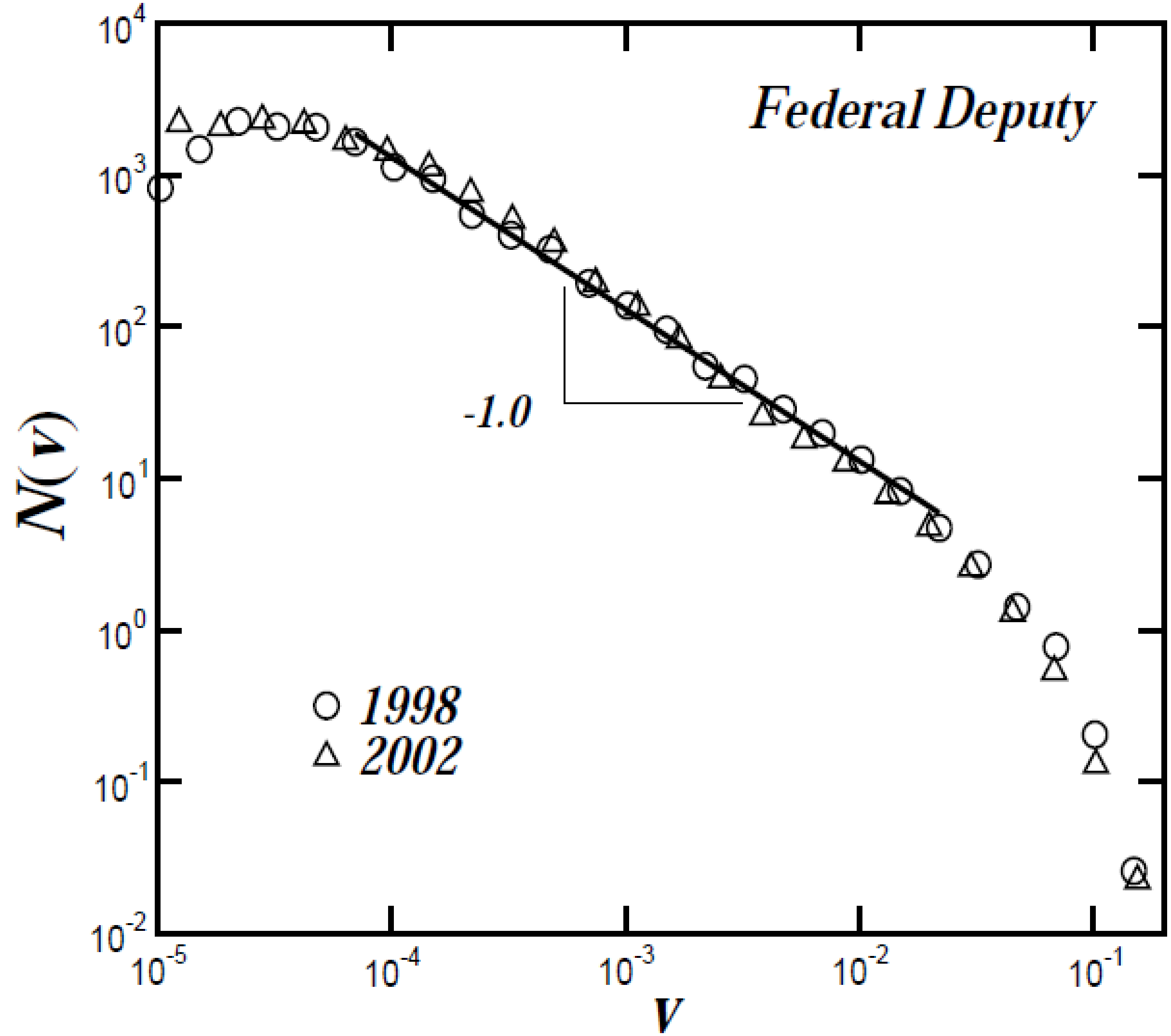
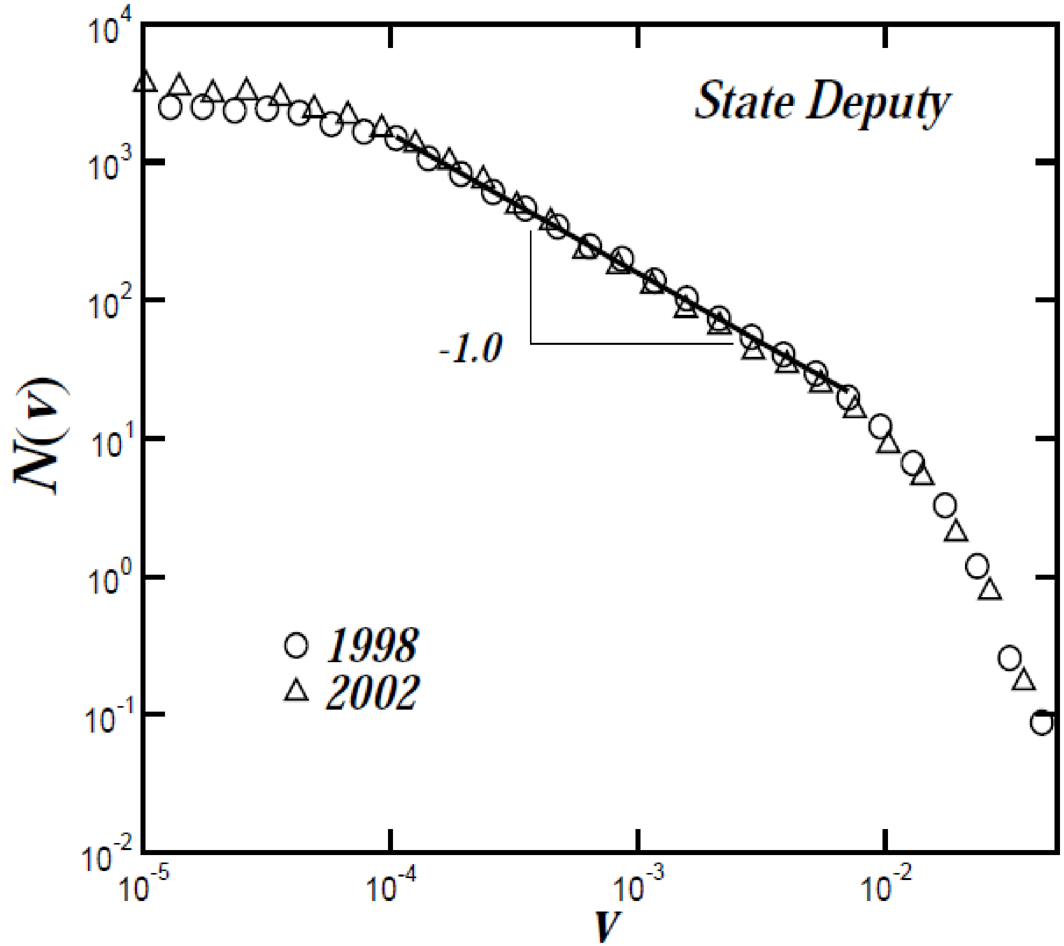


FIG. 2: (a) Double logarithmic plot of the voting distribution for state deputies in 1998 (circ
and 2002 (triangles). The solid lines are the least-squares fits to the data in the scaling regi
The numbers indicate the scaling exponent . (b) The same as in (a) but for federal deputies.



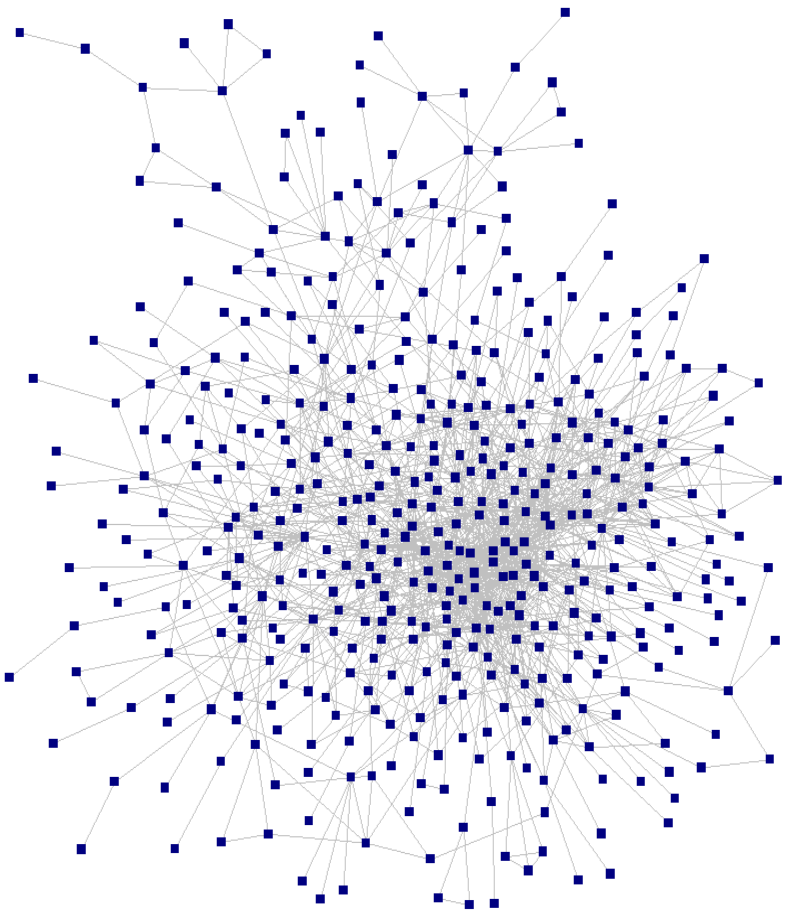
Brazilian elections: voting for a scaling democracy

R. N. Costa Filho*, M. P. Almeida, J. E. Moreira, J. S. Andrade Jr.

Departamento de Física, Universidade Federal do Ceará, Caixa Postal 6030

Scientific Collaboration Network

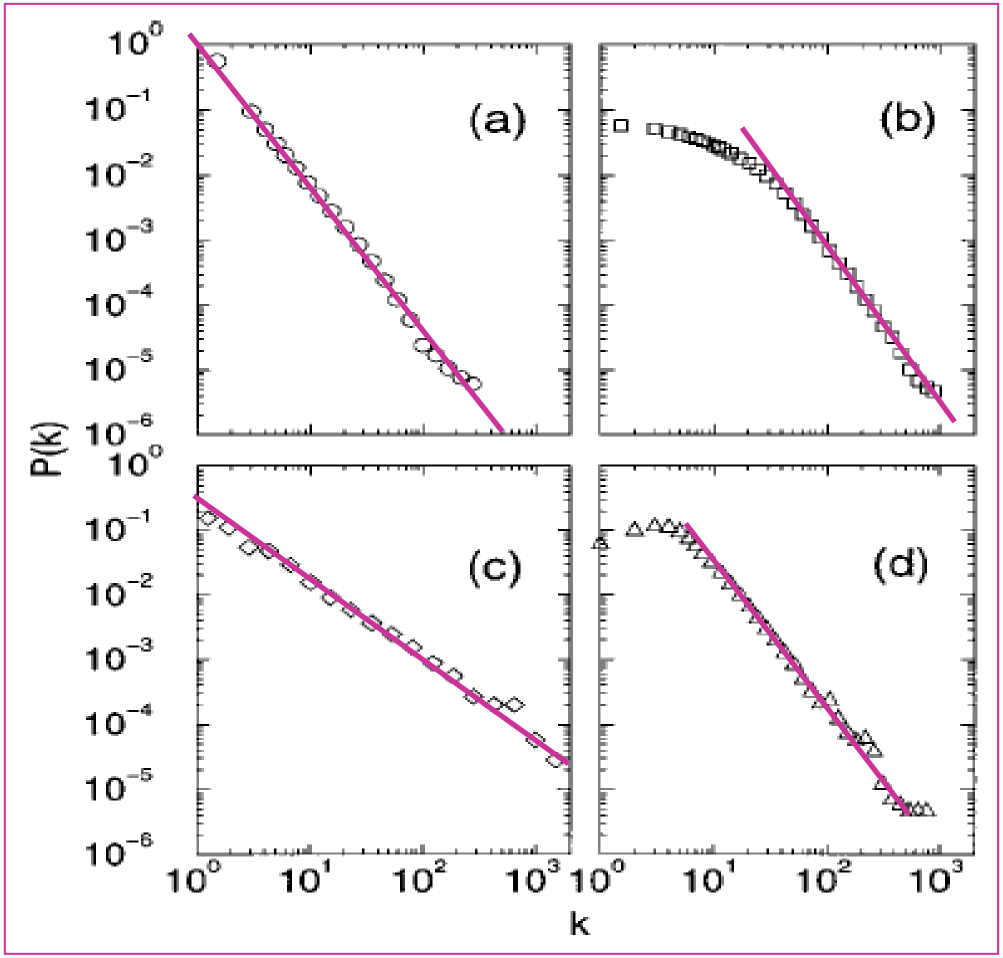
- 400,000 nodes, authors in *Mathematical Reviews* database
- An edge between two authors if they have a joint paper
- Just 676,000 edges



Picture from orgnet.com

Redes Sociais

Albert and Barabasi (1999)

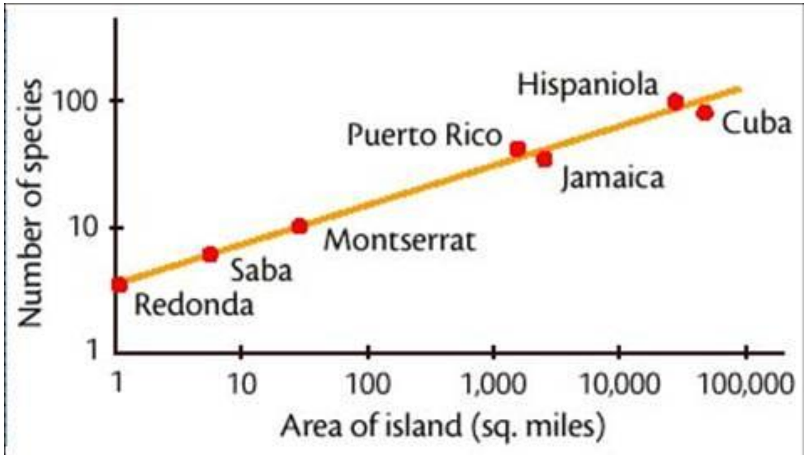


Power laws in real networks:
(a) WWW hyperlinks
(b) co-starring in movies
(c) co-authorship of physicists
(d) co-authorship of neuroscientists

** Same Velfredo Pareto, who defined Pareto optimality in game theory.*

Biogeography and Species Richness

- Number of species on an island is related to its size.
- In general, a 10 fold larger area will have twice the number of species in a given taxa.
- Conservation biologists have used this generalization to predict species loss from habitat destruction and to determine optimum preserve size.

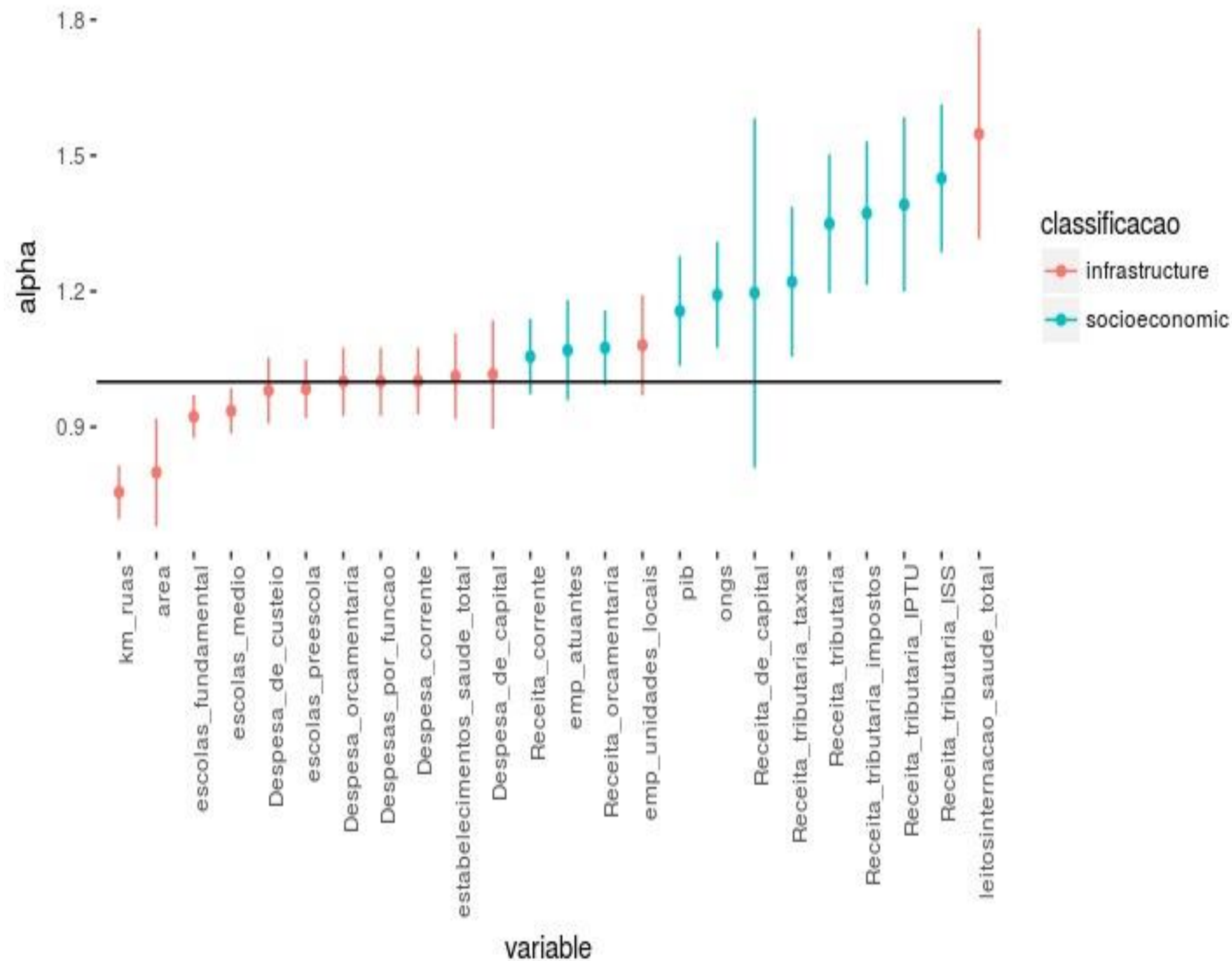


Numbers of species of reptiles and amphibians on Caribbean islands. On a logarithmic plot, the relationship of species number to area is roughly linear.

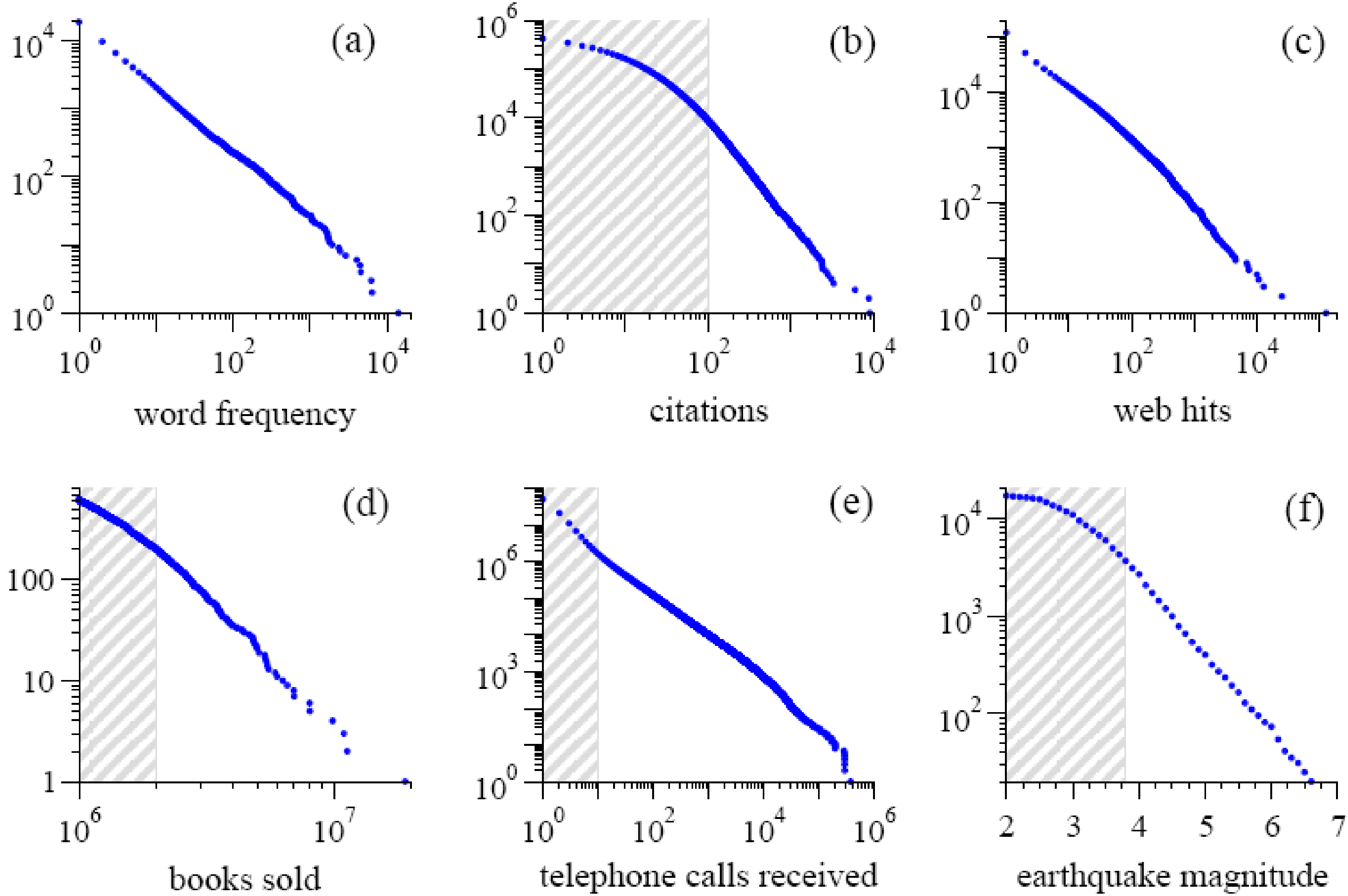
Table 1. Scaling exponents for urban indicators vs. city size

Y	*β*	95% CI	Adj-R^2	Observations	Country–year
New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
"Supercreative" employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
GDP	1.15	[1.06,1.23]	0.96	295	China 2002
GDP	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003
Total housing	1.00	[0.99,1.01]	0.99	316	U.S. 1990
Total employment	1.01	[0.99,1.02]	0.98	331	U.S. 2001
Household electrical consumption	1.00	[0.94,1.06]	0.88	377	Germany 2002
Household electrical consumption	1.05	[0.89,1.22]	0.91	295	China 2002
Household water consumption	1.01	[0.89,1.11]	0.96	295	China 2002
Gasoline stations	0.77	[0.74,0.81]	0.93	318	U.S. 2001
Gasoline sales	0.79	[0.73,0.80]	0.94	318	U.S. 2001
Length of electrical cables	0.87	[0.82,0.92]	0.75	380	Germany 2002
Road surface	0.83	[0.74,0.92]	0.87	29	Germany 2002

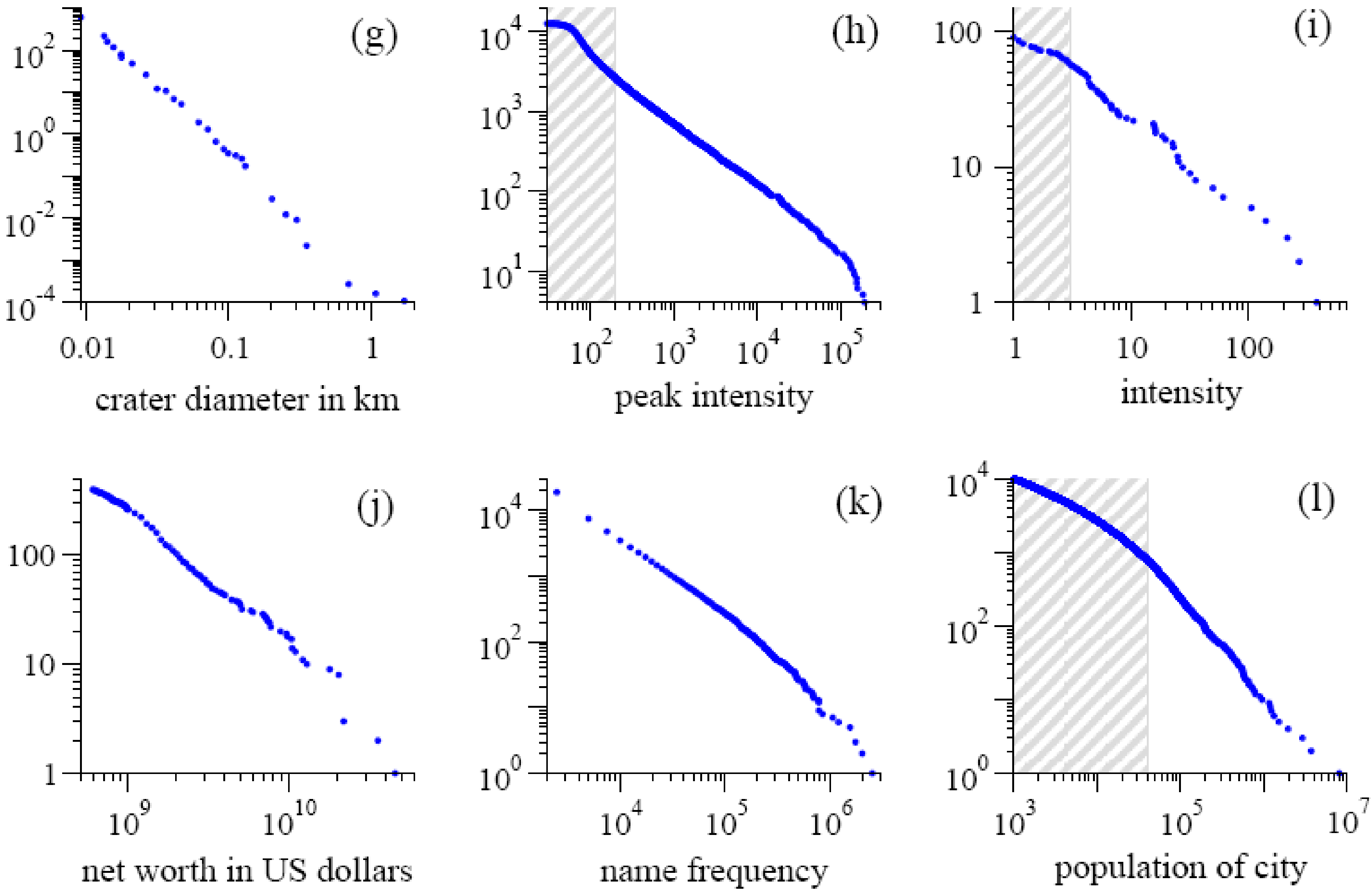
Data sources are shown in *SI Text*. CI, confidence interval; Adj-R^2, adjusted R^2; GDP, gross domestic product.



1.8
1.5
1.2
0.9
alpha
km_ruas
area
escolas_fundamental
escolas_medio
Despesa_de_custeio
escolas_preescola
Despesa_orcamentaria
Despesas_por_funcao
Despesa_corrente
estabelecimentos_saude_total
Despesa_de_capital
Receita_corrente
emp_atuantes
Receita_orcamentaria
emp_unidades_locais
pib
ongs
Receita_de_capital
Receita_tributaria_taxas
Receita_tributaria
Receita_tributaria_impostos
Receita_tributaria_IPTU
Receita_tributaria_ISS
leitosinternacao_saude_total
variable
classificacao
infrastructure
socioeconomic



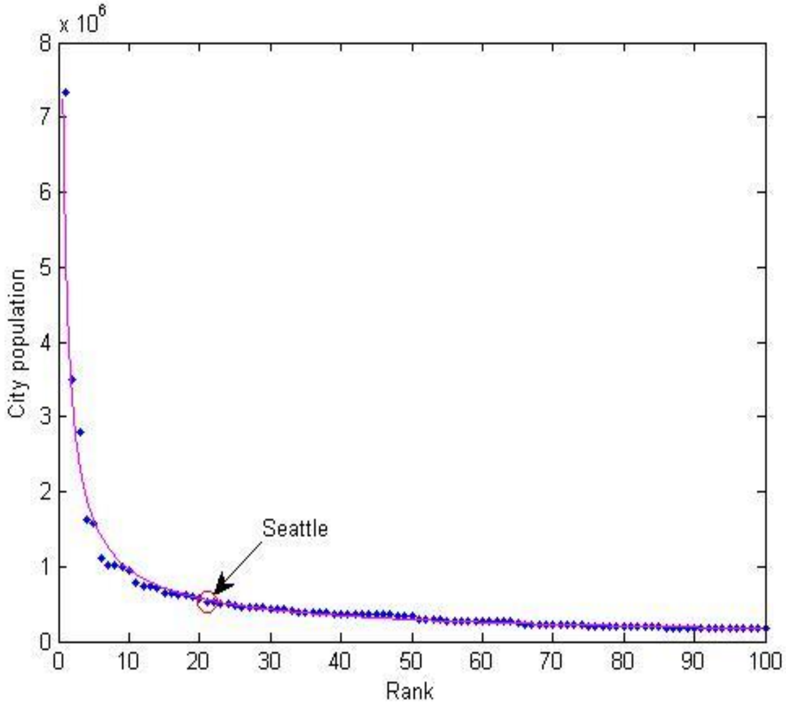
(a)
word frequency
(b)
citations
(c)
web hits
(d)
books sold
(e)
telephone calls received
(f)
earthquake magnitude

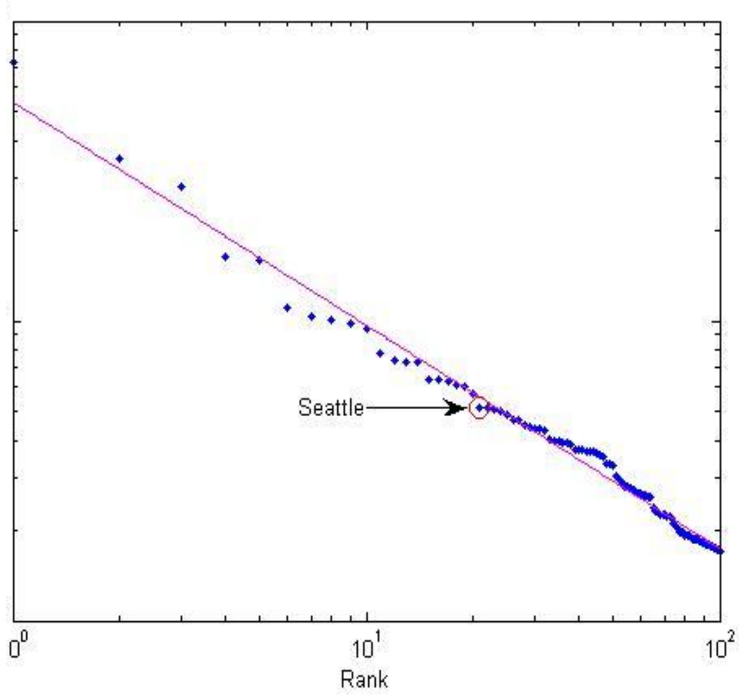


(g)
crater diameter in km
(h)
peak intensity
(i)
intensity
(j)
net worth in US dollars
(k)
name frequency
(l)
population of city

Example: City Populations

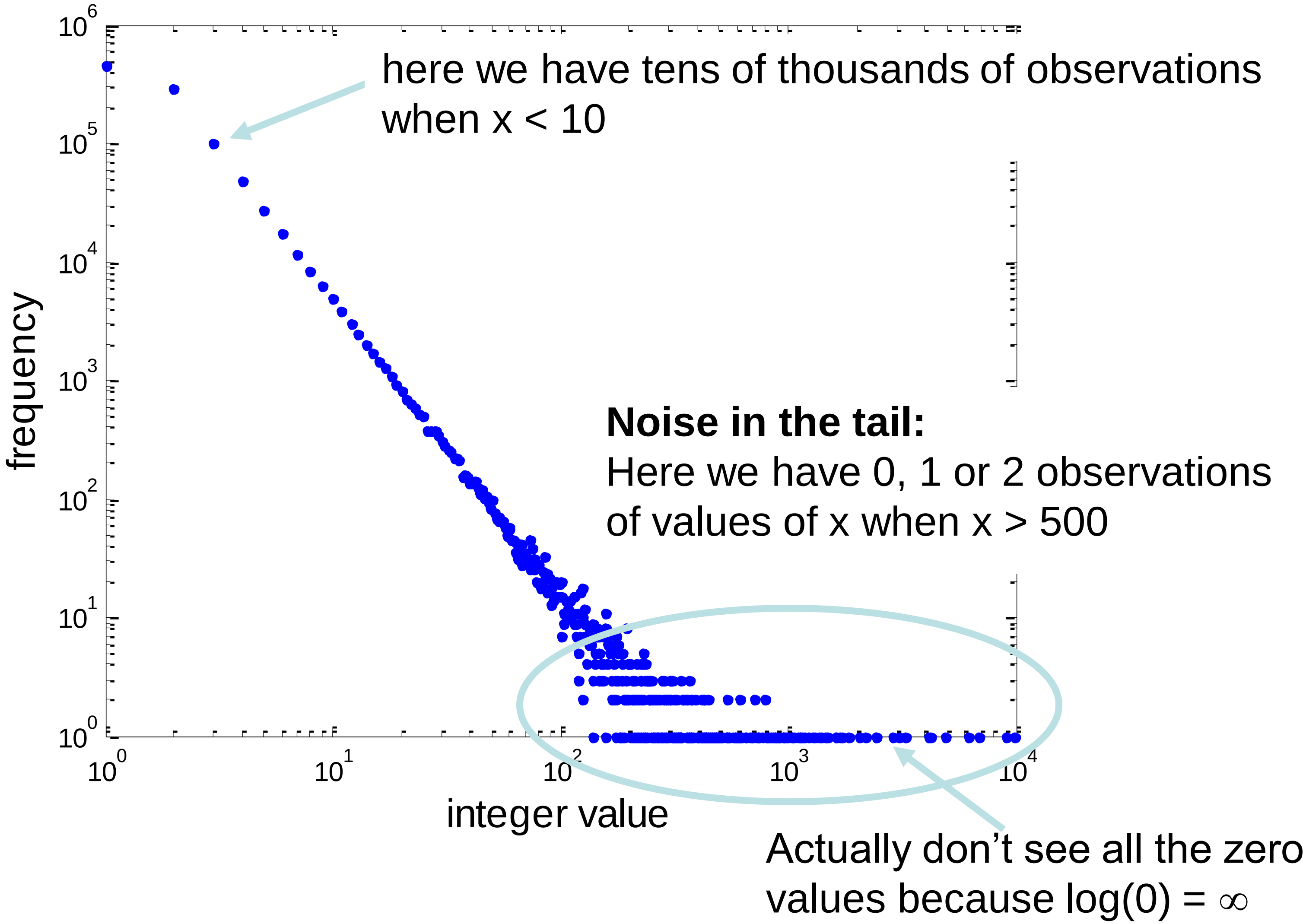
- Power law exponent: c = 0.74





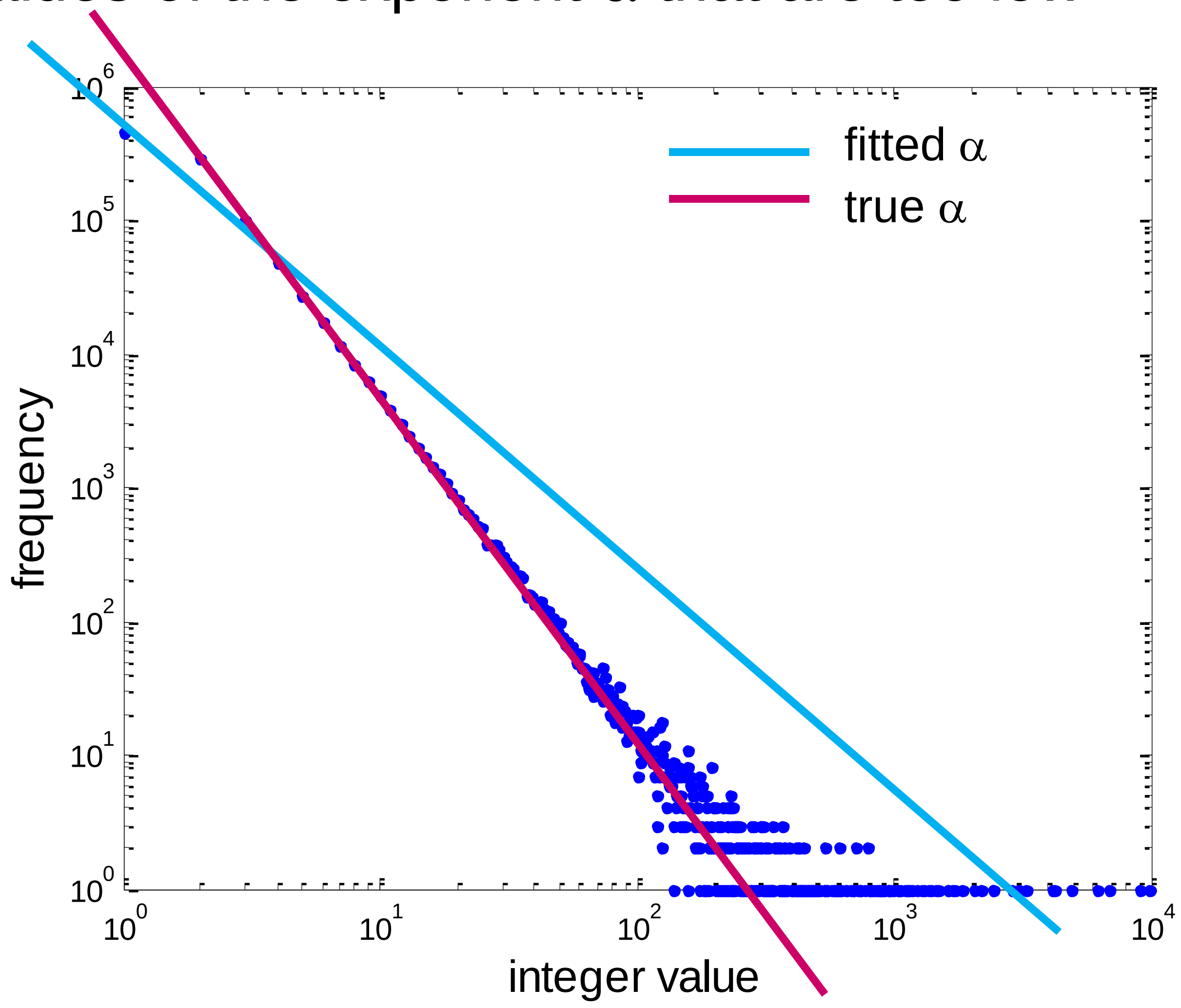
Log-log scale plot of straight binning of the data

- Same bins, but plotted on a log-log scale



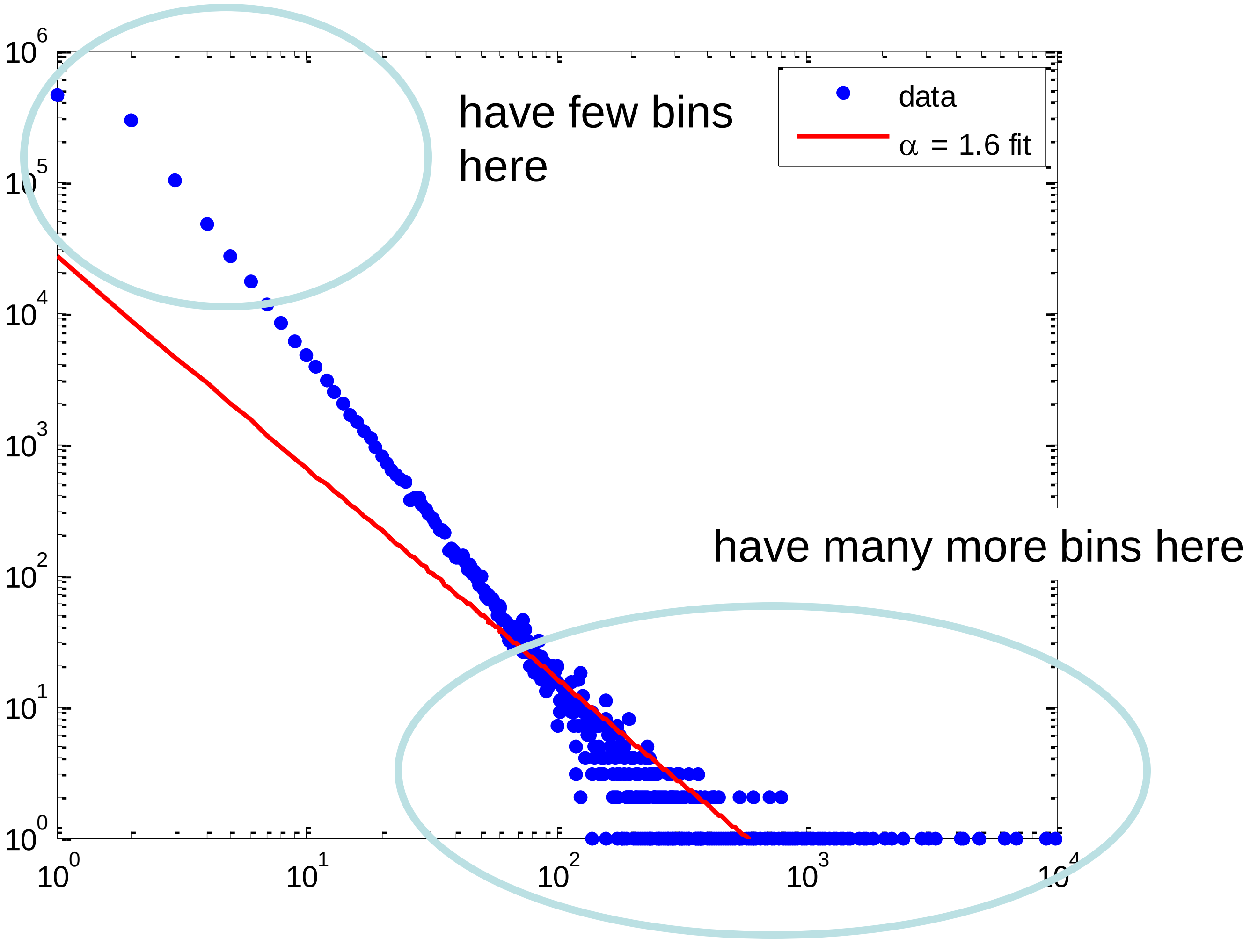
Log-log scale plot of straight binning of the data

- Fitting a straight line to it via least squares regression will give values of the exponent α that are too low



What goes wrong with straightforward binning

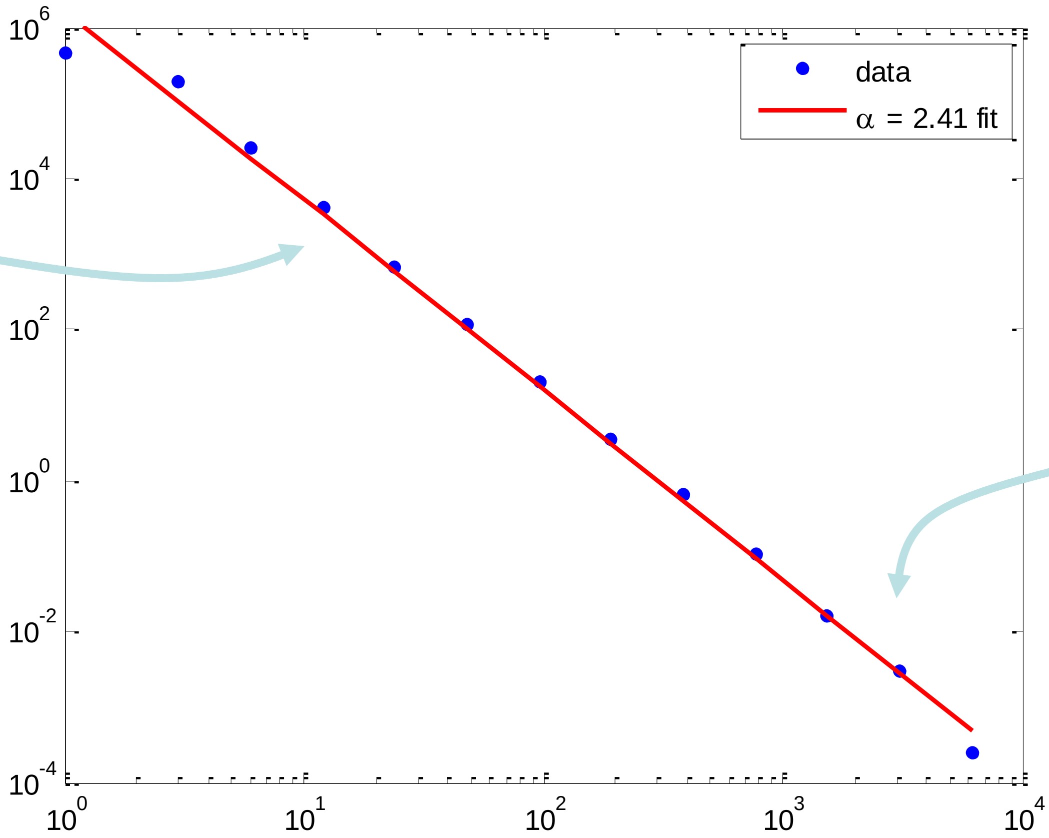
- Noise in the tail skews the regression result



First solution: logarithmic binning

- bin data into exponentially wider bins:
  - 1, 2, 4, 8, 16, 32, …
- normalize by the width of the bin

evenly spaced datapoints



less noise in the tail of the distribution

- disadvantage: binning smoothes out data but also loses information

Second solution: cumulative binning

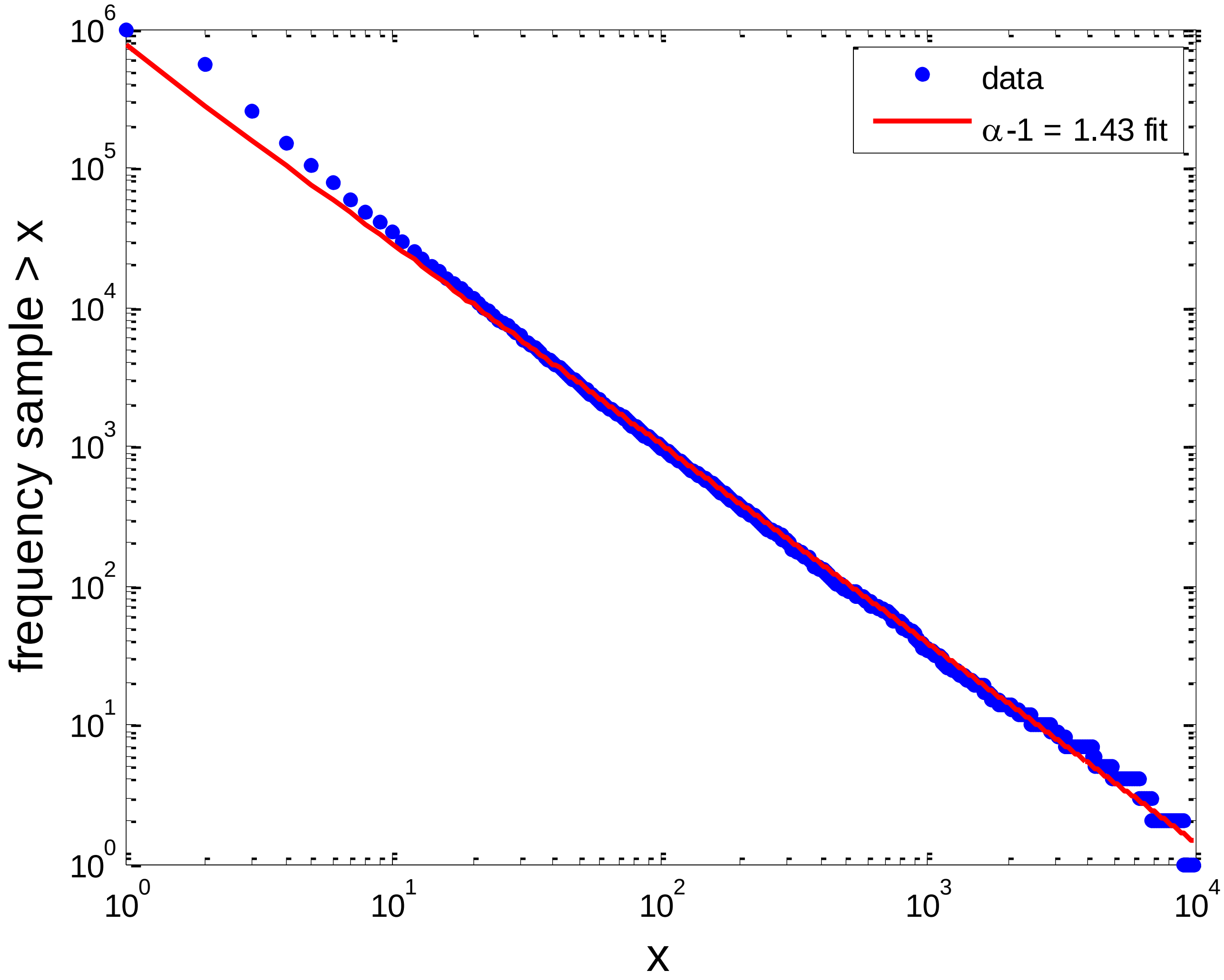
- No loss of information
  - No need to bin, has value at each observed value of x
- But now have cumulative distribution
  - i.e. how many of the values of x are at least X

  - The cumulative probability of a power law probability distribution is also power law but with an exponent
    α - 1

$$\int c x^{-\alpha} = \frac{c}{1-\alpha} x^{-(\alpha-1)}$$

Fitting via regression to the cumulative distribution

- fitted exponent (2.43) much closer to actual (2.5)

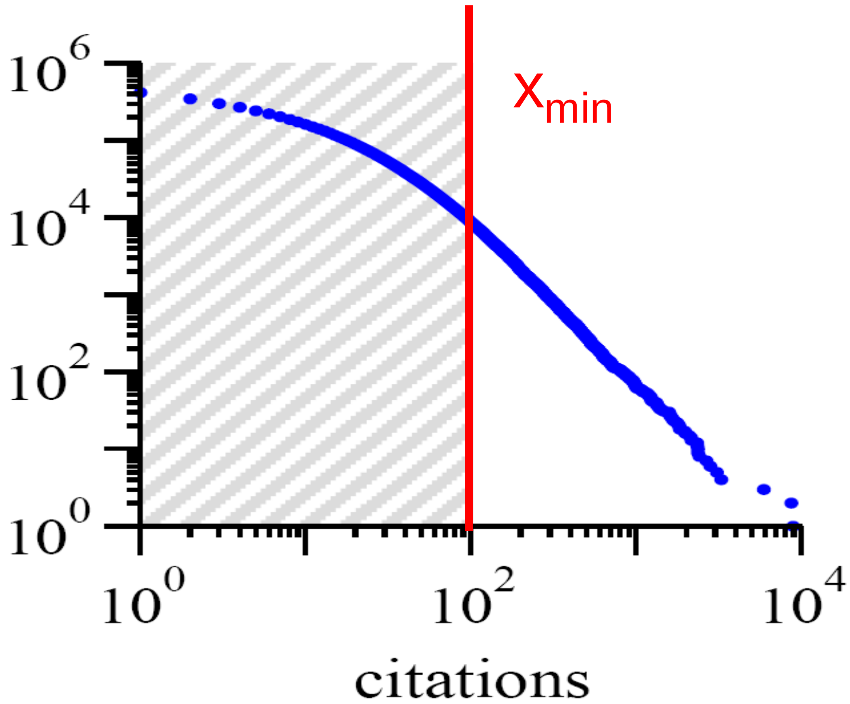


Where to start fitting?

- some data exhibit a power law only in the tail
- after binning or taking the cumulative distribution you can fit to the tail
- so need to select an x_{min} the value of x where you think the power-law starts
- certainly x_{min} needs to be greater than 0, because $x^{-\alpha}$ is infinite at $x = 0$

Example:

- Distribution of citations to papers
- power law is evident only in the tail
  - $x_{min} > 1$



Source: MEJ Newman, 'Power laws, Pareto distributions and Zipf's law

Maximum likelihood fitting – best

- You have to be sure you have a power-law distribution
  - this will just give you an exponent but not a goodness of fit

$$\alpha = 1 + n\left[\sum_{i=1}^{n} \ln \frac{x_i}{x_{\min}}\right]^{-1}$$

- x_i are all your datapoints,
  - there are n of them
- for our data set we get $\alpha = 2.503$ – pretty close!

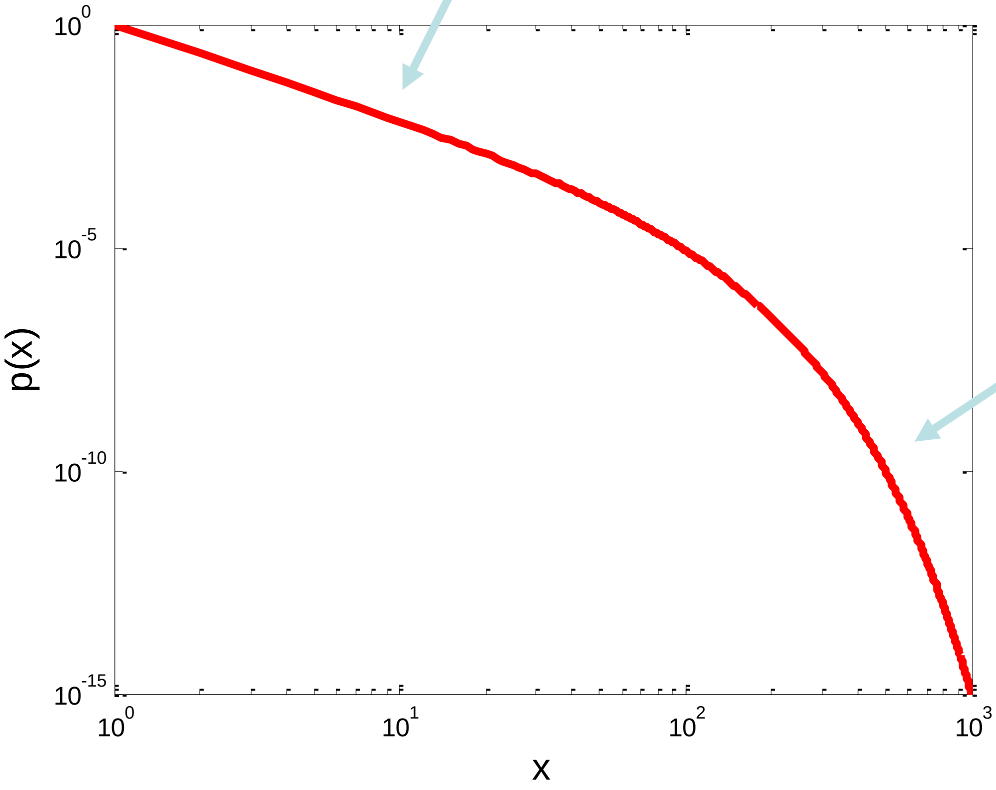
Real world data for x_{min} and α

	x_{min}	α
frequency of use of words	1	2.20
number of citations to papers	100	3.04
number of hits on web sites	1	2.40
copies of books sold in the US	2 000 000	3.51
telephone calls received	10	2.22
magnitude of earthquakes	3.8	3.04
diameter of moon craters	0.01	3.14
intensity of solar flares	200	1.83
intensity of wars	3	1.80
net worth of Americans	$600m	2.09
frequency of family names	10 000	1.94
population of US cities	40 000	2.30

Another common distribution: power-law with an exponential cutoff

- $p(x) \sim x^{-a} e^{-x/\kappa}$

starts out as a power law



ends up as an exponential

but could also be a lognormal or double exponential…

What (universal?) mechanisms give rise to this specific distribution?

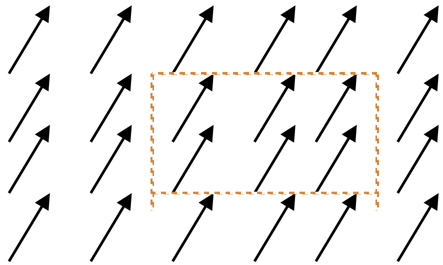
How can we know with rigor when a phenomenon shows PLD behavior?

Mechanismos que geram Leis de Potência

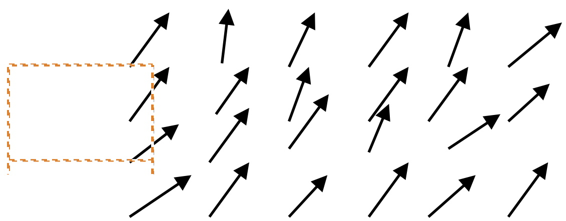
1- Transições de Fase
2- Criticalidade Auto-Organizada (SOC)
3-Fractais
4- Combinação de Exponenciais
5- Processos de Levy
6- Processos de Yule
7- Alometria

Critical phenomena: Phase transitions.

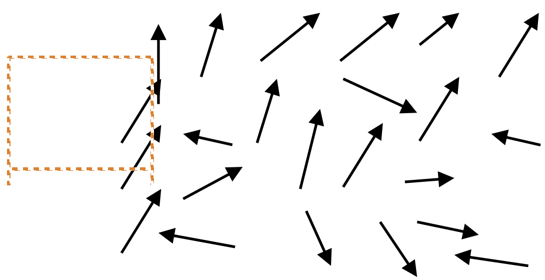
1. T=0
well ordered



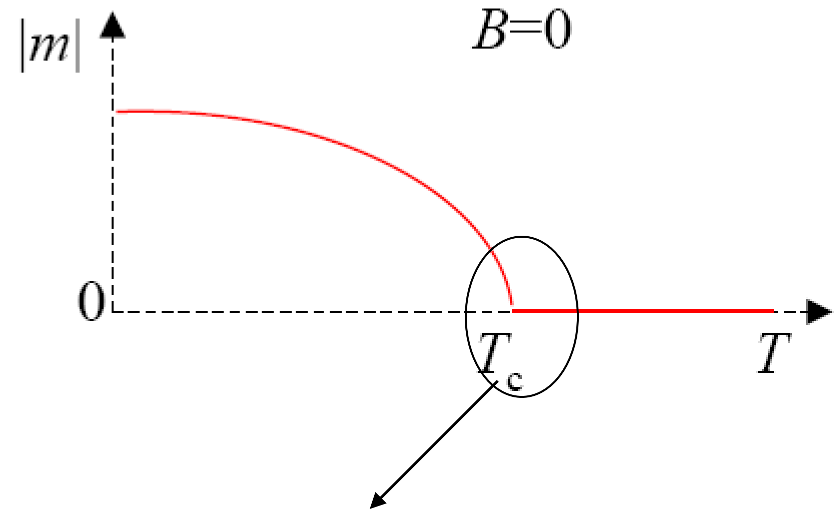
2. 0<T<Tc
ordered



3. T>Tc
disordered



Global magnetization



$c_{\mathrm{H}} \sim |T - T_{\mathrm{c}}|^{-\alpha}$ $\alpha = 0$ heat capacity for $B = 0$

$|m| \sim |T - T_{\mathrm{c}}|^{\beta}$ $\beta = \frac{1}{2}$ coexistence curve

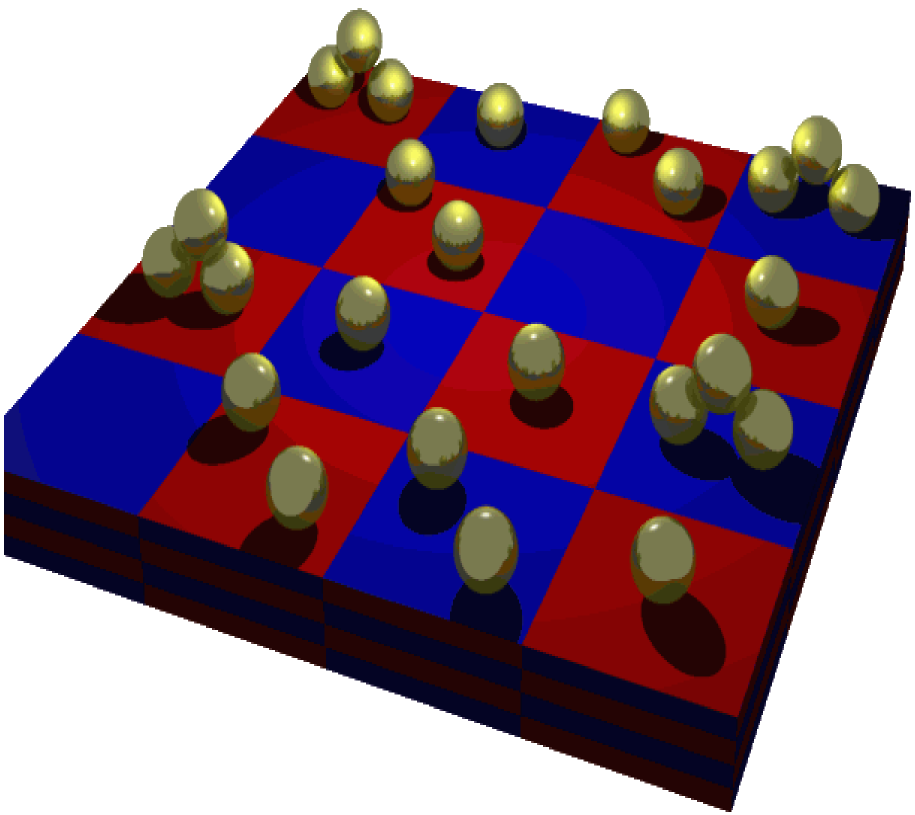
$\chi \sim |T - T_{\mathrm{c}}|^{-\gamma}$ $\gamma = 1$ susceptibility for $B = 0$

$|m| \sim |B|^{\delta}$ $\delta = 3$ magnetization for $T = T_{\mathrm{c}}$

PLD’s

Sandpile model : celular automata

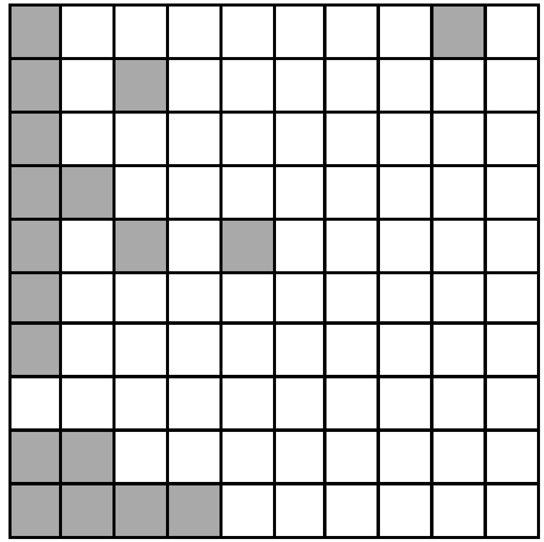
[sandpile applet]

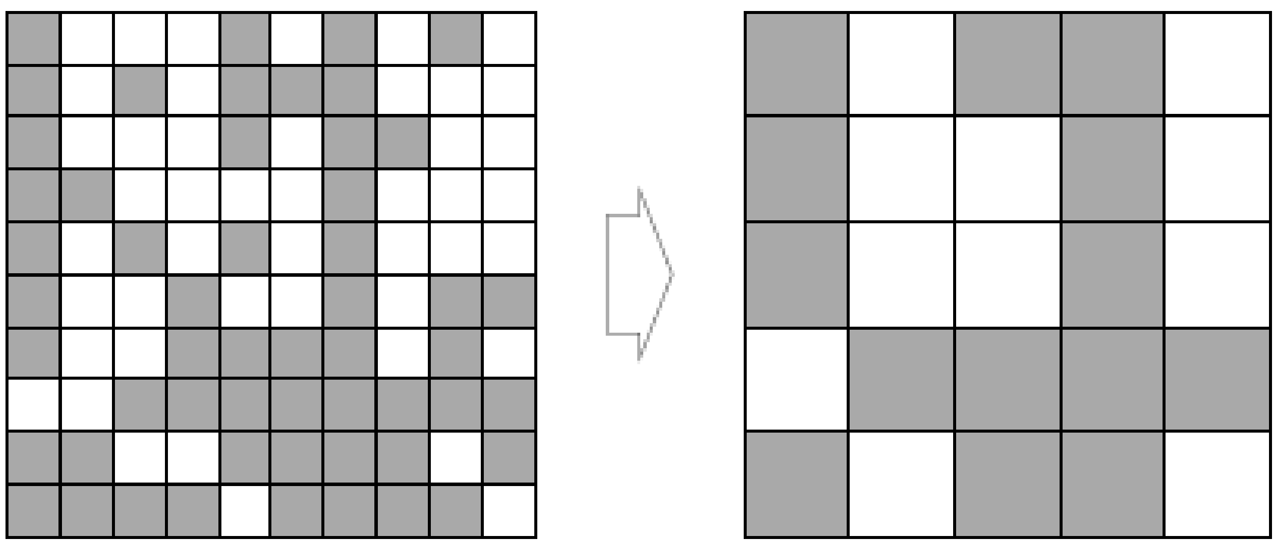


1. A grain of sand is added at a randomly selected site: z(x,y) -> z(x,y)+1;

2. Sand column with a height $z(x,y)>z_c=3$ becomes unstable and collapses by distributing one grain of sand to each of it's four neighbors.
 This in turn may cause some of them to become unstable and collapse (topple) at
 the next time step.
 Sand is lost from the pile at the boundaries. That is why any avalanche of topplings eventually dies out and sandpile "freezes" in a stable configuration with z(x,y)<=z everywhere. At this point it is time to add another grain of sand.

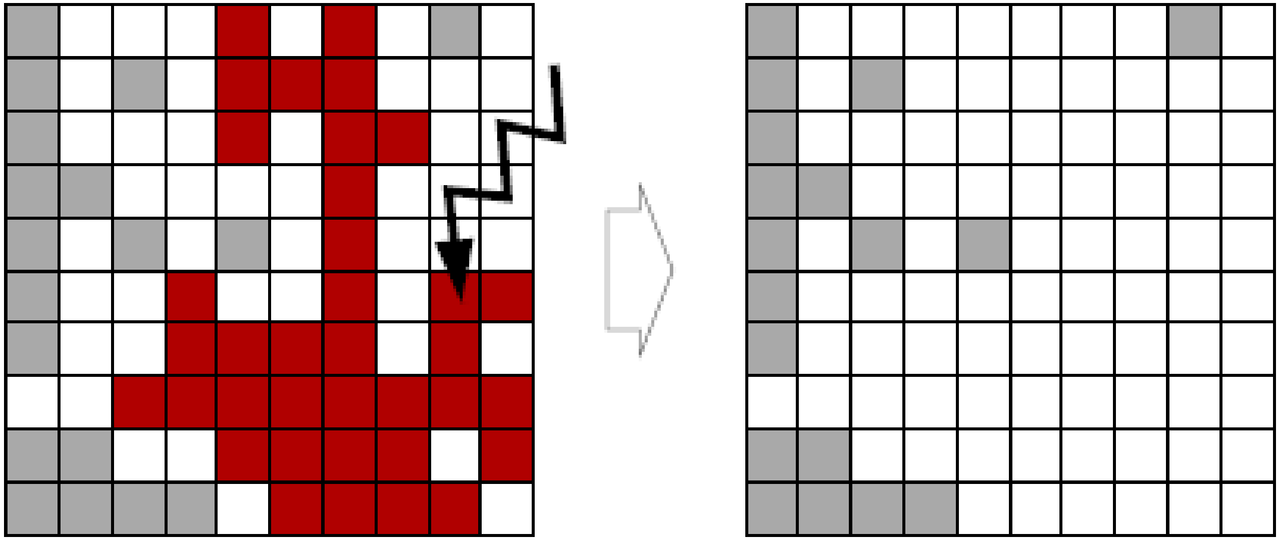
Percolação





SOC: modelo de incêndio em floresta

- Sítios na cor cinza contém árvores
- Sítios na cor vermelha significa árvores em chama
- Sítios vazios não contém árvores
- A cada rodada uma árvore pode nascer em uma célula vizinha a alguma célula com árvore.
- Com baixa probabilidade, uma chama inicia em um sítio aleatório com árvore e pode se propagar ao longo do cluster.



The Yule process (rich gets richer)

Also known as

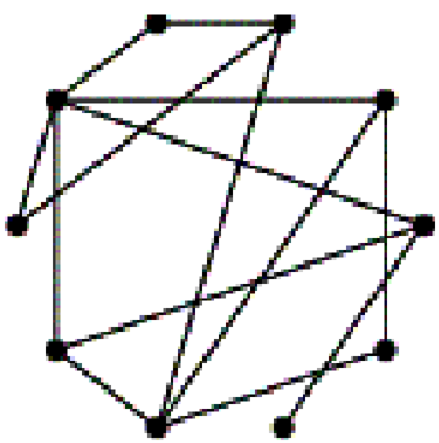
- The gibrat principle (Biometrics)
- Matthew effect
- Cumulative advantage (bibliometrics)
- **Preferential attachment (complex networks)**

- Initial population

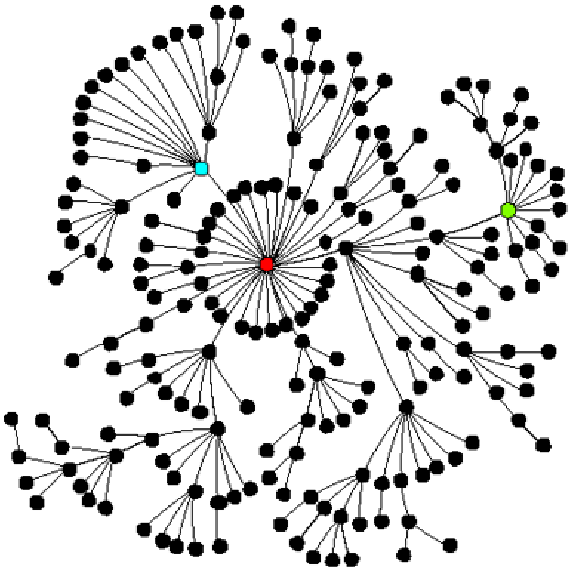
- With t, a new item is added to the population

how?? With probability p, to the most relevant one!
with probability 1-p, randomly.

Initial population



Time (more nodes)



Combinations of exponentials.

Exponential distribution is **more common than PLD**, for instance:

- Survival times for decaying **atomic nuclei**
- **Boltzmann distribution** of energies in statistical mechanics
- etc...

$$p(y) \sim \mathrm{e}^{ay}$$

- Suppose some quantity **y** has an exponential distribution
- Suppose that the quantity we are interested in is **x**, exponentially related to **y**

$$x \sim \mathrm{e}^{by}$$

Where a, b are constants. Then the probability distribution of **x** is a PLD

$$\boxed{\begin{gathered} p(x) = p(y)\frac{\mathrm{d}y}{\mathrm{d}x} \sim \frac{\mathrm{e}^{ay}}{b\mathrm{e}^{by}} = \frac{x^{-1+a/b}}{b} \\ \text{with exponent } \alpha = 1 - a/b \end{gathered}}$$

Log-normal distributions: multiplicative process

- At every time step, a variable N is multiplied by a random variable.
- If we represent this process in **logarithmic space**, we get a **brownian motion**, as long as log(ξ) can be redefined as a random variable.

$$N(t+1) = \xi N(t) \longrightarrow \log N(t+1) = \log(\xi) + \log N(t)$$

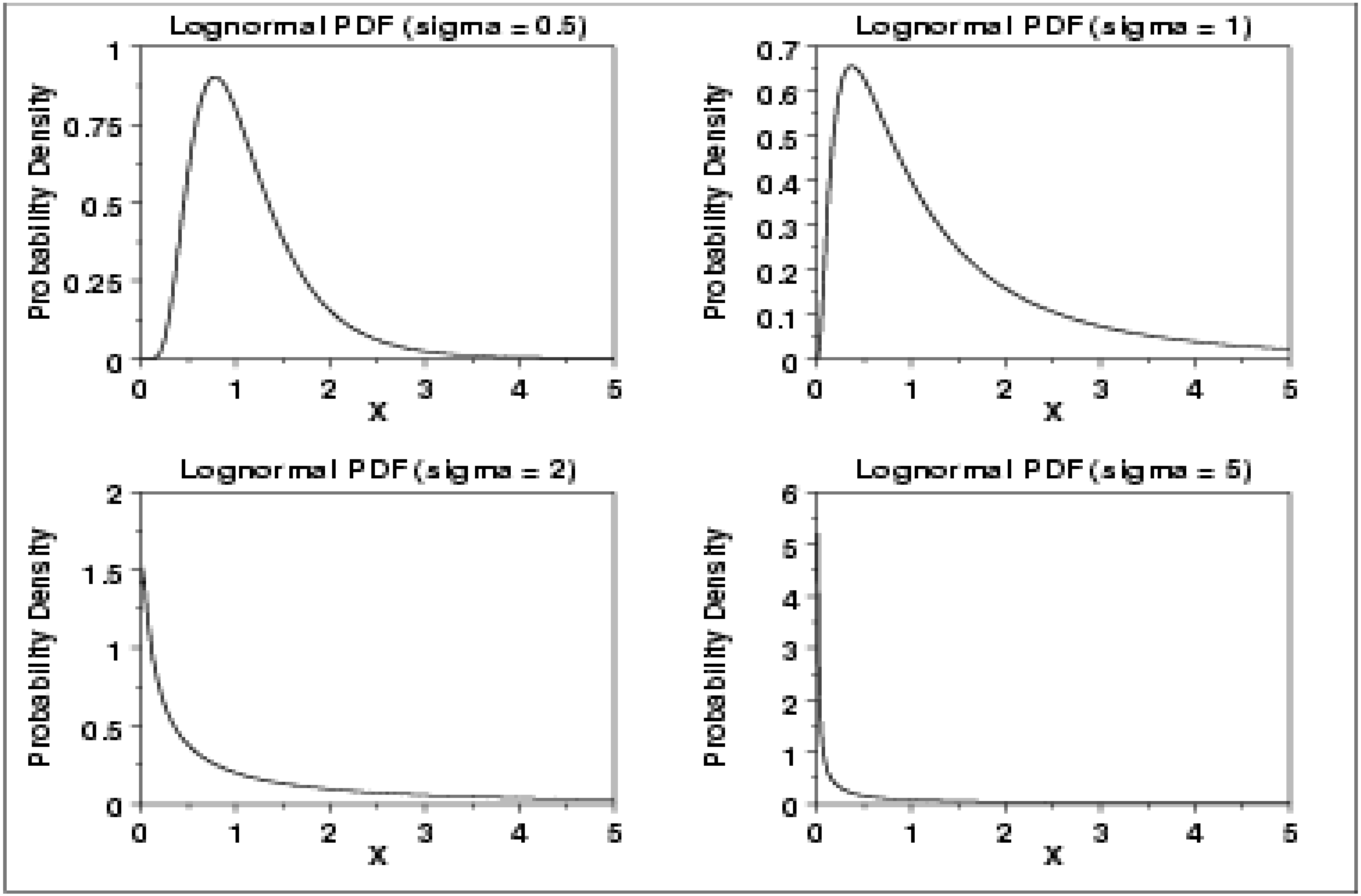
log(N(t)) has a normal (time dependent) distribution (due to the **Central Limit Theorem**)

N(t) is thus a (time dependent) **log-normal** distribution.

Now, **a log-normal distribution looks like a PLD (the tail) when we look at a small portion on log scales** (this is related to the fact that any quadratic curve looks straight if we view a sufficient small portion of it).

A log-normal distribution has a PL tail that gets wider the higher variance it has.

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln x-\mu)^2/2\sigma^2}$$



Example: wealth generation by investment.

•A person invests money in the stock market

• Getting a percentage return on his investsments that varies over time.

• In each period of time, its investment is multiplied by some factor which fluctuates (random and uncorrelatedly) from one period to another.

→ Distribution of wealth: log-normal

Stable Laws: GAUSSIAN and LEVY LAWS

The Lévy laws

Paul Lévy discovered that in addition to the Gaussian law, there exists a large number of stable pdf's. One of their most interesting properties is their asymptotic Power law behavior. Asymptotically, a symmetric Lévy law stands for

$\mathbf{P(x) \sim C / |x|^{1+\mu}}$ for x → infinity

- C is called the tail or scale parameter
- μ is positive for the pdf to be normalizable, and we also have $\mu<2$ because for higher values, the pdf would have finite variance, thus, according to the Central Limit theorem, it wouldn't be stable (convergence to the gaussian law). At this point a generalized central limit theorem can be outlined.

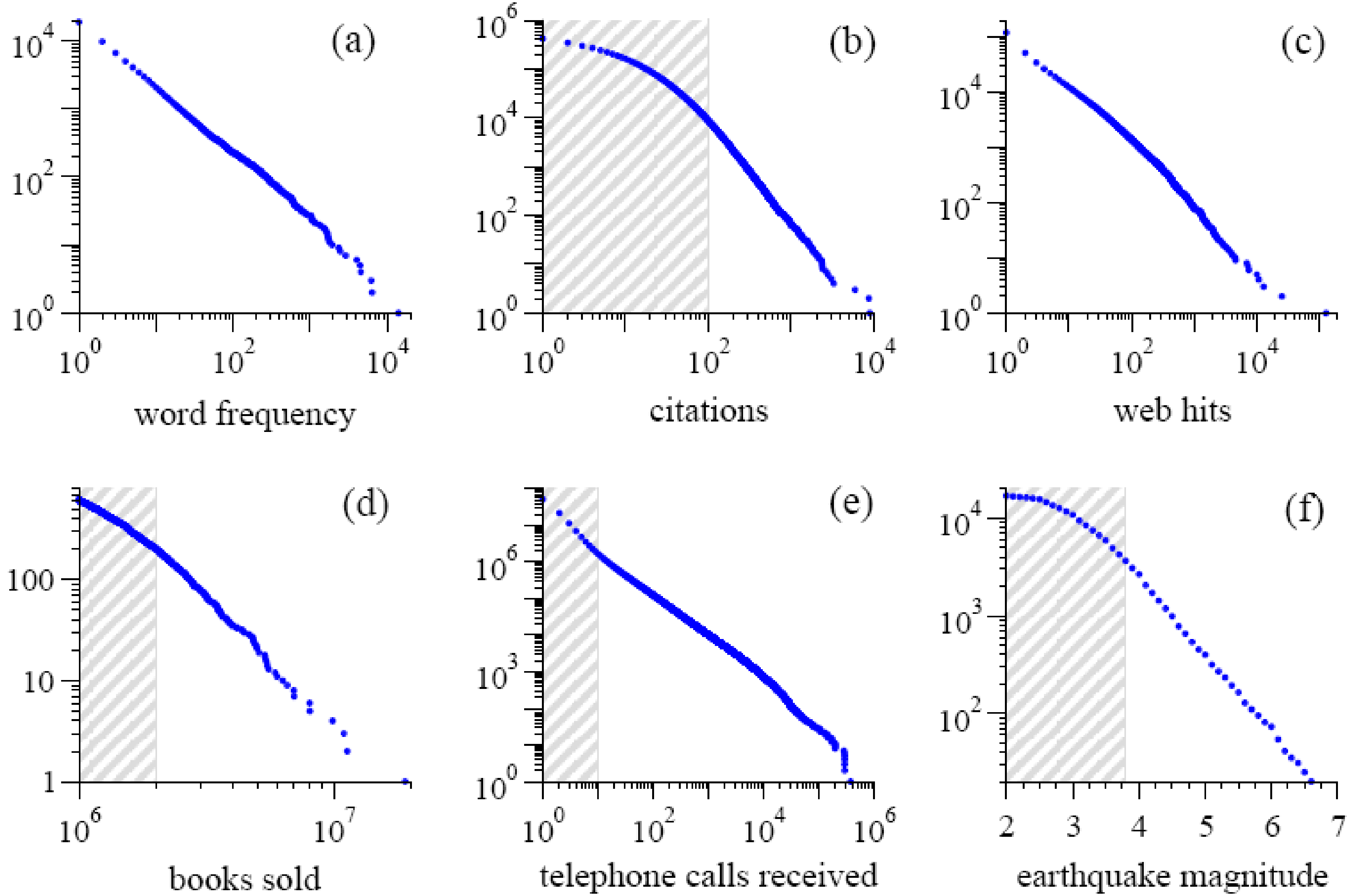
There are not simple analytic expressions of the symmetric Lévy stable laws, denoted by **L** μ**(x)**, except for a few special cases:

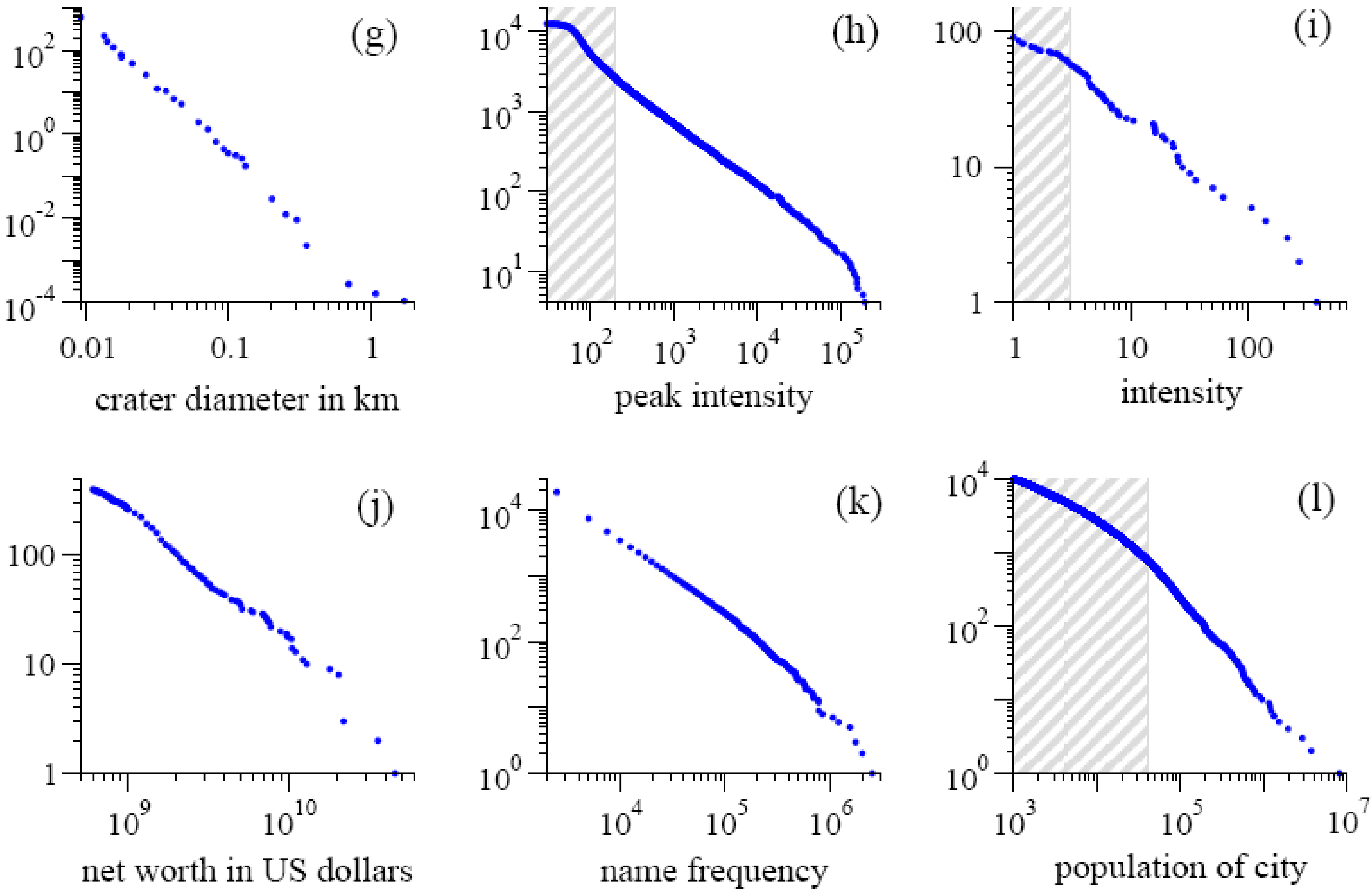
- $\mu=1$ - Cauchy (Lorentz) law - $\mathbf{L1(x) = 1/(x^2 + \pi^2)}$
- $\mu= 1/2$ $\quad f(x;c) = \sqrt{\frac{c}{2\pi}}\ \frac{e^{-c/2x}}{x^{3/2}}$ with C=1

Leis de potência e não-normalidade

- Na natureza parece ser que os eventos raros existem com maior probabilidade do que a normalidade espera.
- ***statistical physics***: critical phenomena, edge of chaos, fractals, SOC, scale-free networks,...
- ***geophysics***: sizes of earthquakes, hurricanes, volcanic eruptions...
- ***astrophysics***: solar flares, meteorite sizes, diameter of moon craters,...
- ***sociology:*** city populations, language words, notes in musical performance, citations of scientific papers...
- ***computer science***: frequency of access to web pages, folder sizes, ...
- ***economics***: distributions of losses and incomes, wealth of richest people,...
- a huge etc.

Ler Black Swan


(a)
(b)
(c)
(d)
(e)
(f)
word frequency
citations
web hits
books sold
telephone calls received
earthquake magnitude


(g)
crater diameter in km
(h)
peak intensity
(i)
intensity
(j)
net worth in US dollars
(k)
name frequency
(l)
population of city

Perigos de Baixo Risco

Desastres naturais produzem consequências que variam em tamanho e frequência.

Perigos de baixo risco são definidos como perigos que historicamente produzem um alto expoente.
Se o expoente é maior do que 1, o risco diminui a medida que a consequência aumenta. De fato, incidentes de grandes consequências são tão raros que sua contribuição para o risco é quase nulo.

Risco Baixo

Resi
Liên
cia

	Asset/Sector	Consequence	Exponent
Low Risk	S&P500 (1974-1999)	$Volatility	3.1-2.7
	Large Fire in Cities	$Loss	2.1
	Airline Accidents	Deaths	1.6
	Tornadoes	Deaths	1.4
	Terrorism	Deaths	1.4
	Floods	Deaths	1.35
	Forest Fires in China	Land Area	1.25
	East/West Power Grid	Megawatts	1
	Earthquakes	Energy, Area	1
	Asteroids	Energy	1
	Pacific Hurricanes	Energy	1

Risco Alto

Baixa

Perigos de Baixo Risco

O expoente é também uma medida de resiliência. Grandes valores de expoente indica grande resiliência. O inverso também se aplica. Baixos valores dele indica baixa resiliência.

Quando o expoente q é menor do que 1 o perigo passa a ser de alto risco. Em outras palavras, o risco aumenta com o aumento da consequência. Incidentes de grandes consequências são mais prováveis para estes perigos, o que resulta em maior risco.

Risco Baixo

Resi Liên cia

	Asset/Sector	Consequence	Exponent
	Hurricanes	$Loss	0.98
	Public Switched Telephone	Customer-Minutes	.91
	Forest Fires	Land Area	.54-.66
	Hurricanes	Deaths	.58
High Risk	Earthquakes	$Loss	.41
	Earthquakes	Deaths	.41
	Wars	Deaths	.41
	Whooping Cough	Deaths	.26
	Measles	Deaths	.26
	Small Fires in Cities	$Loss	.07

Risco Alto

Baixa

Bibliography

- Power laws, Pareto distributions and Zipf´s law,
  M.E.J. Newman
- Critical phenomena in natural sciences,
  D. Sornette
- Problems with Fitting to the PLD
  M. Goldstein, S. Morris, G.G. Yen
- Logarithmic distributions in reliability analysis
  B.K. Jones
- A Brief History of Generative Models for Power Law and Lognormal Distributions
  M. Mitzenmacher